Whole blood mRNA in prostate cancer reveals a 4-gene androgen regulated panel.

Benjamin C. Thomas[1,2], Jonathan D. Kay[1,2,3], Suraj Menon[4, 5], Sarah L.Vowler[4, 5], Sarah N. Dawson[4], Laura J. Bucklow[2], Hayley J. Luxton[2,3], Thomas Johnston[1], Charlie E. Massie[1,6], Michelle Pugh[7], Anne Y. Warren[8], Peter Barker[9],Keith Burling[9], Andy G. Lynch[10], Anne George[1], Johanna Burge[1], Marie Corcoran[1], Sara Stearn[1], Alastair D. Lamb[1], Naomi L. Sharma[1], Greg L. Shaw[1,11], David E. Neal[1,12], Hayley C. Whitaker[1,2,3].

[1]Uro-Oncology Research Group, [2]Biomarker Initiative, [4]Bioinformatics and Statistics Core Facility, [6]Molecular and Computational Diagnostics Group, [7]Genomics Core Facility and [10]Computational Biology Group, Cancer Research UK Cambridge Institute, Robinson Way, Cambridge CB2 0RE, [3]Molecular Diagnostics and Therapeutics Group, University College London, WC1E 6BT, [5]Astra Zeneca, 2 Riverside, Granta Park, Cambridge CB21 6GP, [8]Department of Histopathology and [8]National Institute for Health Research Cambridge Biomedical Research Centre Core Biochemistry Assay Laboratory, Cambridge University Hospitals NHS Foundation Trust, Cambridge Biomedical Campus CB2 2QQ, [11]University College Hospital at Westmoreland Street, 16 - 18 Westmoreland St, London, W1G 8PH, [12]Nuffield Department of Surgical Sciences, John Radcliffe Hospital, Headington, Oxford, OX3 9DU.

Corresponding author:

Dr Hayley Whitaker
Molecular Diagnostics and Therapeutics Group,
Lab 2.4, Cruciform Building
Division of Surgery and Interventional Science

University College London,
WC1E 6BT.
Tel: +44 (0)2076799170 Fax: +44 (0)2031082305

29 Email: Hayley.Whitaker@ucl.ac.uk

30

31 Short title: Whole blood mRNA in prostate cancer

32

33 Key words: PAXgene, prostate cancer, androgen receptor, biomarker

34

35 Word count: 5650

36

37 None of the authors have any conflict of interest to report in association with this work.

Abstract

Due to increased sensitivity the expression of circulating nucleotides is rapidly gaining popularity in cancer diagnosis. Whole blood mRNA has been used in studies on a number of cancers, most notably two separate studies that used whole blood mRNA to define non-overlapping signatures of prostate cancer that has become castration independent. Prostate cancer is known to rely on androgens for initial growth and there is increasing evidence of the importance of the androgen axis in advanced disease. Using whole blood mRNA samples from patients with prostate cancer we have identified the 4-gene panel of *FAM129A*, *MME*, *KRT7* and *SOD2* in circulating mRNA that are differentially expressed in a discovery cohort of metastatic samples. Validation of these genes at the mRNA and protein level was undertaken in additional cohorts defined by risk of relapse following surgery and hormone status. All four genes were down-regulated at the mRNA level in the circulation and in primary tissue but this was not always reflected in tissue protein expression. *MME* demonstrated significant differences in the hormone cohorts while *FAM129A* is down-regulated at the mRNA level but is raised at the protein level in tumours. Using published chIP-seq data we have demonstrated that this may be due to AR binding at the *FAM129A* and *MME* loci in multiple cell lines. These data suggest that whole blood mRNA of androgen regulated genes has the potential to be used for diagnosis and monitoring of prostate cancer.

Introduction

Prostate cancer is the leading cause of cancer death in men (Statistics 2013). Prostate cancer growth relies upon androgens, such as testosterone, binding to the androgen receptor (AR) and promoting AR binding to DNA via androgen response elements where it regulates the transcription of androgen regulated genes (Claessens, et al. 2014). Due to the importance of the AR signalling axis, treatment for prostate cancer relies upon blocking the production of androgens or their binding the AR. Although this treatment is effective, patients will eventually relapse as prostate cancer cells continue to grow, despite continued hormone therapy. This stage of disease was originally termed hormone independent or refractory but recent evidence, including the use of drugs to block adrenal androgen synthesis (abiraterone) or binding to the AR (enzalutamide), suggests that the AR axis remains active in advanced prostate cancer and is a valid therapeutic target (Mateo, et al. 2014). This is supported by the critical involvement of androgen regulated genes to predict aggressive prostate cancer in a recently published gene signature found in tissue (Klein, et al. 2014).

Diagnostic prostate cancer biomarkers primarily detect differences between benign prostates and those containing primary tumour. Despite the large number of published manuscripts describing biomarkers for detecting prostate cancer, the detection of androgen regulated prostate specific antigen (PSA) in serum remains the primary biochemical test used for diagnosis (Egger, et al. 2013; Kattan, et al. 1998; Partin, et al. 1993). PSA is not an ideal prostate cancer biomarker as it can be elevated by a number of benign conditions including benign prostatic hyperplasia and prostatitis. PSA has a sensitivity of approximately 80% but specificity as low as 20% resulting in false positive and negative results (Catalona, et al. 1994; Thompson, et al. 2005). *PCA3*, a small non-coding RNA that can be detected in urine, is increasingly being used for prostate cancer diagnosis, particularly in patients with raised PSA and at least one negative biopsy (de Kok, et al. 2002; Haese, et al. 2008). This is despite the requirement for prostatic massage and reports of relatively poor sensitivity (65%) and specificity (60%) (Hessels and Schalken 2009). Unlike *PCA3*, PSA has shown some promise in predicting aggressive disease and relapse following treatment but this remains an area of prostate research where

84 novel markers could significantly impact on clinical decisions (Augustin, et al. 2013; Botchorishvili,
85 et al. 2009).
86
87 Over 30 years ago, Leon et al. showed that circulating DNA levels were increased in cancer patients
88 compared to healthy controls (Leon, et al. 1977). Circulating nucleic acids as biomarkers have
89 advantages over proteins such as the ability to be amplified and detected with high sensitivity and
90 specificity. Expression arrays and real-time quantitative PCR (qPCR) allow quantification of many
91 genes in a single experiment. (Schwarzenbach, et al. 2011). More recently whole blood mRNA has
92 been investigated for the discovery and development of biomarkers as RNA is more labile and may
93 more accurately reflect any early changes in cells leading to tumour development (Miura, et al. 2005;
94 Papadopoulou, et al. 2006; Williams 2010). The PAXgene system is used for the storage and
95 purification of RNA from blood and provides storage of blood samples for 50 months at -20^{o}C
96 (Rainen, et al. 2002). Its use has enabled the investigation of differences between whole blood mRNA
97 expression levels in patient samples with haematological and rheumatological diseases, breast, thyroid
98 and prostate cancers (Batliwalla, et al. 2005; Lewis, et al. 2011; Li, et al. 2004; Yang, et al. 2011).
99 In prostate cancer there have been a number of recent publications utilising PAXgene samples for
100 prognosis to identify those primary prostate cancers that are likely to be life limiting (indolent primary
101 versus aggressive disease) (Danila, et al. 2013; Liong, et al. 2012; Marin-Aguilera, et al. 2015; Olmos,
102 et al. 2012; Ross, et al. 2012). In back-to-back publications Olmos et al. and Ross et al. showed different
103 gene signatures, both capable of predicting shortened survival in patients with castration-resistant
104 prostate cancer (Olmos et al. 2012; Ross et al. 2012). Using genes pre-selected by literature review Ross
105 et al. used unamplified whole blood mRNA to define a six-gene model which separated castrate-
106 resistant patients into a low risk group with median survival of 34.9 months compared to a high risk
107 group with a median survival of 7.8 months (Ross et al. 2012). Using amplified whole blood mRNA
108 Olmos et al. used Affymetrix expression arrays and Bayesian latent process decomposition to identify
109 a signature which divided their cohort into 4 groups with survival ranging from 9.2 months to 25.6
110 months (Olmos et al. 2012). That these two manuscripts showed no overlap between the genes in their

signature for castration-resistant disease may reflect the use of candidate genes and RNA amplification which is known to introduce bias in expression arrays in the Olmos, but not the Ross, manuscript (Kitchen, et al. 2011). Other groups have used RNA from urine or the peripheral circulation to study previously identified and well characterised markers for prostate cancer using PCR-based methods and found they were robust, particularly when used in conjunction with circulating tumour cells (Danila et al. 2013; Quek, et al. 2012).

Here we describe the identification of a whole blood mRNA gene signature from unamplified PAXgene samples using Illumina HT-12 expression arrays. Genes were identified using the extremes of prostate pathology; a metastatic and benign discovery cohort. Differentially expressed genes were further validated in additional primary cohorts with defined hormonal status or risk stratification. In addition the gene panel was investigated at both the gene and protein level in the circulation and primary prostate tissue and AR binding to the gene locus determined.

Materials and methods

Unless otherwise stated all reagents were purchased from Sigma Aldrich (St. Louis, USA).

Study design

An overview of the study design is given in Figure 1A. All PAXgene samples were taken from patients enrolled in the ProMPT trial at Addenbrookes Hospital, Cambridge (National Institute of Health and Research ID 5837) between 2010 and 2012. Informed consent was obtained from each patient after full explanation of the purpose and nature of all procedures used.

The discovery cohort used PAXgene samples from 22 patients; 12 control patients with an elevated PSA and negative TRUS biopsy (benign), and 10 patients with metastatic prostate cancer. These patients also had core biopsy specimens taken from the primary prostate available for immunohistochemistry. An additional 10 samples from men with raised PSA who had undergone

template biopsy were analysed for comparison with the benign control group. The clinical characteristics are given in Table 1.

The gene signature was validated in 2 cohorts; a defined hormone status cohort and a defined risk cohort. Hormone status was determined as those patients not receiving any hormone treatment (hormone naïve), patients receiving hormone treatment and responding (stable disease) or patients no longer responding to hormone therapy (hormone refractory). For the risk cohort patients were defined as low, intermediate or high risk based on the following clinical criteria; low risk – PSA ≤10ng/ml, Gleason 6, ≥T2, intermediate risk – PSA ≥10ng/ml but ≤20ng/ml, Gleason 7, ≥T2 or high risk – PSA ≥20ng/ml, Gleason ≥8, ≥T3. The clinical characteristics of the risk cohort are given in Table 2. For this study hormone refractory patients are defined as those with two consecutive PSA rises to ≥0.2ng/ml. The clinical characteristics are given in Table 3.

RNA extraction and cDNA formation

2.5ml of whole blood was taken from patients using PAXgene tubes (GenProbe, (San Diego, USA)) and stored according to the manufacturer's instructions. RNA was extracted using the PAXgene RNA Blood kit (Qiagen, (Limburg, Netherlands)) and eluted in 80μL of elution buffer. RNA quantification was performed by absorbance (OD A260nm) on the Nanodrop ND1000 instrument (Thermo Scientific, (Cambridge, UK)). 500ng of RNA per sample was reverse transcribed to cDNA using High Capacity RNA-to-cDNA Master Mix (Applied Biosystems, (California, USA)).

RNA Expression Array

Using 22 samples in the discovery cohort (benign (n=12) and metastatic n=10)), gene expression analysis was carried out on Illumina Human HT12 version 4 arrays (Illumina, (San Diego, USA)). Data analyses were carried out on R using Bioconductor packages (Gentleman 2004). Raw intensity data from the array scanner was processed using the BASH and HULK algorithms as implemented in the *beadarray* package (Cairns 2008; Dunning 2007). Log2 transformation and quantile normalisation of the data was performed across all sample groups. Differential expression analysis was carried out using

the *limma* package (Smyth 2005). Genes were selected for further analysis by ranking according to an unadjusted p-value and a cut off of 0.01 applied to stratify genes more likely to be significantly altered. Following interrogation of the top ranked genes data was ranked by p-value and then fold change. Further stratification of genes was completed using published datasets to identify differentially expressed genes associated with poor prognosis(Glinsky, et al. 2004; Taylor, et al. 2010). Genes with published antibodies for immunohistochemistry or good differential expression between tumour and benign glands in Human Protein Atlas were preferentially selected. A table of how genes were selected and what criteria were taken into account is shown in Supplementary Table 2.

A false discovery rate (FDR) of 0.39 was calculated for the top 100 probes indicating that only 61/100 were likely to be true positives and that additional validation was required to confirm any candidate genes.

Quantitative Real-Time Polymerase Chain Reaction

Quantitative real-time Polymerase Chain Reaction (qPCR) was performed in triplicate using 2pmol Sigma primers, 5µl of SYBR Green and 10ng of cDNA template using the Applied Biosystems 7900HT Real-Time PCR system. Primer sequences are given in Supplementary Table 3. Ct values were calculated and the expression of target genes was normalised against the expression of RPLP2 housekeeping gene using the δδCt method. RPLP2 was chosen as a housekeeping gene by analysing all genes in the expression array with high expression and minimal variation (Supplementary Figure 1).

Chromatin Immunoprecipitation:

Chromatin was immunoprecipitated from LNCaP cells treated with either 1nM R1881 or vehicle (ethanol) for 1hour as described in Massie et al.(Massie and Mills 2009) using 10µg AR n-20 antibody (sc-816, Santa Cruz Biotechnology) and an equal mixture of Dynabeads Protein A and G (10001D and 10003D, Life Technologies). Primer sequences are shown in Supplementary Table 4. Results are shown relative to the vehicle control. P-values were calculated using a Wilcoxon rank-sum test.

Immunohistochemistry

For immunohistochemistry tissue cores taken from the primary prostate of patients within the discovery cohort were used. In addition a TMA was constructed from multiple benign and tumour samples from 104 patients who underwent radical prostatectomy as previously described (Whitaker, et al. 2010; Whitaker, et al. 2013). Immunohistochemistry was also performed on a hormone relapsed TMA comprised of 74 patients, of which 9 were on hormone treatment and responding and 50 initially responded to treatment but subsequently underwent biochemical or clinical relapse as described (Ramos-Montoya, et al. 2014). The clinical characteristics are given in Table 4.

All immunohistochemistry was performed using a Bondmax Autostainer (Leica (Solms, Germany)) using the following antibodies; FAM129A (1:750, Atlas Antibodies (Stockholm, Sweden)), KRT7 (2.8g/L, Novocastra (Solms, Germany)), SOD2 (1:1500, Atlas Antibodies) and MME (1:50, Novocastra). Citrate buffer pH6.0 (ER1) was used for all antigen retrieval. KRT7 has been validated and is sold for routine use in the clinical setting using the same conditions and Bondmax Autostainer (Addenbrookes Hospital). The specificity of antibody staining by FAM129A, SOD2 and MME was validated using qPCR and immunohistochemistry of formalin fixed paraffin embedded LNCaP cell pellets transiently transfected with siRNA against each target or a non-targeting control (Supplementary Figure 2). Loss of staining in the specific cellular compartment in up to 50% of cells in the siRNA group was considered specific. Confirmation of pathology was conducted by a uropathologist (AW) for all samples. All scoring was done independently by two observers (one an independent specialist uro-oncology pathologist) who were both blinded to the TMA plan. Staining was classified into the following categories: none (-), weak (+), moderate (++) and high (+++), based on intensity. For the TMA, where multiple pathologies were present in a single core, each was scored independently and multiple scores given for distinct grades of cancer. For core biopsies where cancer was present, it was scored independently of surrounding benign tissue and a single score given for each patient.

ELISA

All assays were performed on the MesoScale Discovery (MSD) platform (Maryland USA). Normal bind plates were coated at 4°C overnight with 30µl goat anti-MME antibody (1.44µg/ml, R&D Systems) before washing and blocking with MSD blocker A. After diluting 1:5 in Delfia diluent II (Perkin Elmer (Massachusetts, USA)), 50µl of serum sample or standard was added to each well and incubated for 2 hours at room temperature. After washing, 25µl of biotinylated goat anti-MME antibody (1:100, R&D Systems (Minneapolis, USA)) was added per well and incubated at room temperature for 1 hour. Signal was detected using streptavidin sulphoTAG (1:1000, MSD) incubated for 30 mins at room temperature. Wells were washed and 150µl MSD read buffer added prior to detection on a Sector 6000 plate reader. Standards ranging from 39-40,000pg/ml were made using recombinant MME (R&D Systems) and Delfia diluent II. A standard curve was run alongside all samples to ensure reproducibility of the assay.

Androgen regulation of LNCaP cells and AR chIP-seq

Gene expression data were generated as part of a previously published study (Massie, et al. 2011). Briefly, LNCaP cells grown for 72 h in steroid depleted medium prior to treatment with the synthetic androgen R1881. Samples included 3 time zero samples; 10 vehicle (ethanol) control samples taken at 2, 4, 8, 12 and 24 h in duplicate and 36 androgen (R1881)-treated samples taken every 30 min for 4 h then every hour until 24 h following treatment (with replicates at 1, 2, 4, 8, 12, 16, 20 and 24 h). Gene expression data was analysed as before (Massie et al. 2011). The AR chIP-seq experiments have previously been described (Massie et al. 2011; Wang, et al. 2009; Yu, et al. 2010).

Data analysis and modelling

Grouped p-values were calculated using a Kruskal-Wallis test. Pairwise analysis was performed using a Mann-Whitney two-tailed t-test. For both analysis a p-value of <0.05 was considered to be statistically significant. Previously published expression array data from prostate tissue was analysed using recursive partitioning to predict time to biochemical recurrence from 179 patients (29 benign, 131 localised prostate cancer and 19 metastatic disease) (Taylor et al. 2010). Recursive partitioning was also

used to determine qPCR cut-offs between groups in the defined hormone status cohort and the defined risk cohort using the ctree function in R (Hothorn 2006). Models were fitted for all 4 genes and serum PSA individually.

Results

Identification of genes capable of discriminating between benign and metastatic patients

To discover if any genes demonstrated significant differential expression between benign and tumour cohorts a discovery cohort were selected which represented the extremes of prostate disease; 10 metastatic and 12 benign samples. To ensure that there were no fundamental differences between the blood samples from each group baseline haematological parameters such as haemoglobin and white blood cells were assessed and were not significantly different (Supplementary Figure 3).

The mRNA from the discovery cohort samples were assessed using Illumina Human HT12 v4 expression arrays with and without haemoglobin reduction, which has been reported to influence measurement of accurate expression using gene arrays (Fan and Hegde 2005; Vartanian, et al. 2009). Although the signal to noise ratio was much lower in the haemoglobin reduced samples, when candidate genes were validated by qPCR using the original PAXgene samples without haemoglobin reduction, they were not significant suggesting that haemoglobin mRNA reduction had altered the samples significantly (Supplementary Figure1A and data not shown). The 100 top differentially expressed genes in the non-haemoglobin reduced samples are shown in Supplementary Figure 4. Six genes were taken forward for further analysis; family with sequence similarity 129, member A (*FAM129A*) also known as *NIBAN*, pyrophosphatase (inorganic) 1 (*PPA1*), dolichyl-phosphate mannosyltransferase polypeptide 3 (*DPM3*), superoxide dismutase-2 (*SOD2*), keratin-7 (*KRT7*) and membrane metallo-endopeptidase (*MME*) also known as *CD10*. Although unrelated, all of these genes have been previously associated with cancer and *DPM3*, *SOD2*, *KRT7* and *MME* have previously been linked to prostate cancer (Supplementary Table 1). The genes identified in previous PAXgene studies failed to cluster our discovery cohort into benign and metastatic samples, probably

as a result of different sample processing methods (data not shown) (Olmos et al. 2012; Ross et al. 2012).

Figure 1B shows a heatmap of microarray data of probes for the six genes that were amongst the top 1000 most differentially expressed probes. Boxplots of these probes are shown in Figure 1C for the benign and metastatic groups, along with the p-values from *limma* tests for differential expression between the groups. Further validation of these results using qPCR of fresh cDNA from same patients showed that three genes, *KRT7*, *DPM3* and *PPA1*, were not significantly different between benign and metastatic groups despite *DPM3* previously being linked to prostate cancer cell invasion (Figure 1D) (Manos, et al. 2001). This result was not unexpected given the high FDR seen in the array probably due to the wide clinical parameters in the discovery cohort, non-specific binding to probes on the expression array or the increased sensitivity of qPCR. As *KRT7* is routinely used as a basal cell marker in prostate cancer we took the decision to allow it to remain in the selected gene panel (Supplementary Table 1) (Ashida, et al. 2006; Manos et al. 2001). The remaining 3 genes *FAM129A*, *SOD2* and *MME* all achieved statistical significance ($p<0.05$). The benign group within the discovery cohort all had raised PSA and at least one negative TRUS biopsy and therefore it was possible that up to 30% may go on to have a later diagnosis of prostate cancer (Shariat and Roehrborn 2008). To establish that this benign cohort was a valid control group qPCR for *FAM129A*, *SOD2*, *MME* and *KRT7* was compared between the TRUS confirmed benign group and a cohort of benign prostates confirmed by template biopsy with a much lower probability of having cancer (<10%). This showed no significant differences and confirmed the benign group as a valid control (Supplementary Figure 5).

Circulating mRNA may reflect gene expression and protein levels in tumour tissue

It is hypothesised that the mRNA collected in PAXgene samples could be influenced by expression of specific genes in primary tumour tissue or be altered due to the immune response (Marin-Aguilera et al. 2015; Olmos et al. 2012; Ross et al. 2012). As no tissue RNA was available from our discovery cohort we examined gene expression of the four genes identified in a published independent cohort (Taylor et al. 2010). Expression of all four genes, *FAM129A*, *SOD2*, *KRT7* and *MME* was

significantly reduced in metastatic samples when compared to benign, consistent with the results we found in circulating mRNA (Figure 2A). To establish if changes in circulating mRNA gene expression and tissue mRNA expression accurately reflected protein expression in the tumour the diagnostic biopsies from the benign prostates or primary tumours of the discovery cohort were examined using immunohistochemistry (IHC) for FAM129A, SOD2, MME and KRT7 (Figure 2B). FAM129A showed cytoplasmic staining and expression was reduced, albeit not significantly (p=0.17) in metastatic biopsy tissue when compared to benign. SOD2 showed ubiquitous punctate staining, consistent with its known mitochondrial localisation (Wispe, et al. 1989). In contrast to the mRNA expression and published data, protein levels increased in metastatic samples compared to benign, although not significantly (p=0.070) (Bostwick, et al. 2000). Both KRT7 and MME appeared to be associated with the plasma membrane, consistent with previous publications (Glass and Fuchs 1988; Zheng, et al. 2010). Protein expression of both decreased significantly with the metastatic group consistent with the whole blood mRNA (Figure 1). However, detection of MME protein in serum was not consistently reduced when measured by ELISA using samples from the same patients (Figure 2C).

Localised disease stratified by risk

Patients with localised prostate cancer that may be destined for surgery can be stratified into risk groups for relapse using a number of criteria include PSA, stage and grade (D'Amico, et al. 1999) Using recursive partitioning of the risk cohort serum PSA could distinguish low and high groups with a ≥11.3ng/ml cut off. We stratified 47 patients into low, intermediate and high risk groups as defined in the materials and methods, and tested their whole blood mRNA levels of *FAM129A*, *KRT7*, *SOD2* and *MME* (Figure 3A). Only circulating *KRT7* and *MME* mRNA demonstrated any significant ability to discriminate between the risk groups. *KRT7* was able to discriminate between low and intermediate risk groups (p=0.0061) and this was confirmed using recursive partitioning where nodes represent the point at which the most significant differences between the two arms exist (cut off ≥0.00042) (Figure 3B). Overall PSA >11.3ng/ml performs better than KRT7 but with a PSA ≤11.3ng/ml they perform similarly. *KRT7* and *MME* could both differentiate between intermediate and high risk

groups (p=0.0019 and p=0.011 respectively). This result was consistent with significant decreases in tissue expression of *KRT7* and *MME* in localised prostate cancer from previously published data (Figure 3C, p<0.0001 and p=0.013 respectively) (Taylor et al. 2010). In the same dataset *FAM129*A exhibited no change between benign and localised tissue groups while *SOD2* expression was raised controls (p=0.018) consistent with our findings at the protein level in the primary tissue biopsies (Figure 2B). Low expression of *FAM129A*, *SOD2* and *MME* mRNA were able to predict poor prognosis in Taylor et al. with multiple probes (Taylor et al. 2010) while low tissue mRNA expression of *FAM129A*, *KRT7* and *SOD2* was also linked to poor outcome in data published by Glinsky et al. (Supplementary Figure 6) (Glinsky et al. 2004).

To establish if changes in tissue mRNA and circulating mRNA in our PAXgene samples reflected protein expression in primary prostate tissue IHC was performed using an in-house Cambridge TMA for FAM129A, KRT7, SOD2 and MME protein (Figure 3D). All four proteins showed a significant difference between the benign and Gleason groups (p<0.0001). In contrast to circulating mRNA expression and primary tissue biopsies from metastatic patients, protein levels of FAM129A showed a significant increase in localised tumours compared to benign tissue. KRT7, SOD2 and MME all exhibited a decrease in protein expression when compared to benign tissue. This decrease was particularly striking for KRT7 and MME.

Androgen regulation of the 4-gene panel

Within the discovery cohort the metastatic patients were receiving a range of treatments which may have influenced the circulating mRNA levels of different genes. We used a well-defined hormone sensitivity cohort to determine the effect of hormone treatments on the circulating mRNA levels of the 4-gene panel of *FAM129A*, *SOD2*, *KRT7* and *MME* (clinical data is given in Table 3). These patients had either never received hormone treatment (hormone naïve), were receiving hormone treatment and responding (hormonal therapy) or were no longer responding to hormone therapy (hormone relapsed). Using a Kruskal Wallis test, detection of the mRNA coding for all 4 genes was significantly different between all groups for each gene (p<0.05) (Figure 4A). Circulating expression of mRNA coding for

FAM129A, *SOD2* and *MME* showed a significant reduction in those patients responding to hormones compared to those receiving no treatment which may indicate androgen regulation of those genes. To determine cut-off points associated with hormone status recursive partitioning of the qPCR data was performed (Figure 4B). This was consistent with recursive partitioning of *FAM129A*, *SOD2* and *MME* demonstrating that the hormone naïve (HN) group could be distinguished from the hormone treated groups (cut-offs >0.337 (*FAM129A*), 1.162 (*SOD2*) and 0.054 (*MME*)). None of the four genes were able to distinguish hormone relapsed patients from those responding to treatment using a Mann Whitney test, however even in the small numbers of patients in these cohorts recursive partitioning of the data for *MME* showed some ability to distinguish patients on hormonal therapy (HT, node 2) from hormone relapsed (HR, nodes 4 and 5) (Figure 4A and 4B).

To establish if these changes in circulating mRNA reflected protein expression in a defined hormone cohort we performed IHC for FAM129A, KRT7, SOD2 and MME on a hormone relapsed TMA comprised of patients who were responding to hormone treatment hormonal therapy (HT)) or relapsed following hormone treatment (hormone relapsed (HR)) (Table 4). These results were compared to the untreated hormone naïve patients used in the in-house Cambridge TMA (Figure 3D). KRT7, SOD2 and MME, but not FAM129A all showed highly significant differences across the three groups. There were no statistically significant pairwise analyses for either FAM129A or KRT7 at the protein level. However, both SOD2 and MME showed significant differences at the protein level when patients were treated with hormone (hormone naïve vs on hormonal therapy) ($p<0.0001$ (SOD2) and $p=0.0028$ (MME))

To determine if these observed changes in circulating mRNA and tissue protein were a result of androgen regulation of the 4-gene panel we examined data from an androgen time course of the LNCaP cell line which showed that only *FAM129A* expression was decreased following androgen treatment, consistent with our data (Figure 5A and Supplementary Figure 7). To determine if the AR binds to potential binding sites in the promoters of the 4-gene panel we examined published chIP-seq data from a variety of cell lines (Massie et al. 2011; Wang et al. 2009; Yu et al. 2010). These results

showed convincing AR binding in multiple prostate cancer cell lines for *FAM129A* and *MME* (Figure 5B). There was evidence for AR binding at the *KRT7* locus, but only in the two *ERG* positive VCaP cell line studies suggesting possible *ERG* involvement in *KRT7* regulation (Massie et al. 2011; Yu et al. 2010) (Supplementary Figure 7). AR binding to the *SOD2* locus was seen in two cell lines (LNCaP and VCaP) but only in data produced in our laboratory (Massie et al. 2011). To confirm AR binding at the promoters of these genes chromatin immunoprecipiation (chIP) was performed in the LNCaP cell line using an anti-AR antibody (Figure 5C). Following 1 hour of treatment with the synthetic androgen, R1881, AR binding was significantly enriched at the FAM129A, KRT7 and MME promoters, but not the SOD2 promoter.

Discussion

The majority of diagnostic biomarkers are identified through studies comparing benign and primary prostate cancer. All previous studies using PAXgene samples to determine circulating mRNA signatures in prostate cancer have examined primary indolent versus aggressive prostate cancer either using pre-selected genes of interest or amplified the mRNA which is known to introduce bias (Danila et al. 2013; Kitchen et al. 2011; Marin-Aguilera et al. 2015; Olmos et al. 2012; Ross et al. 2012). In this study we have described the identification of genes initially using a benign versus metastatic comparison using Illumina Human HT12 v4 expression arrays without amplification or use of haemoglobin reduction which has been reported to improve data quality in PAXgene studies (Fan and Hegde 2005; Vartanian et al. 2009). Although we found that haemoglobin reduction did improve the data quality as defined by Illumina (Supplementary Figure 1), all subsequent qPCR validation in non-haemoglobin reduced samples demonstrated that the mRNA in haemoglobin reduced samples did not accurately reflect the mRNA in the biological sample.

Using expression arrays and native samples we compared the two extremes of prostate pathology in the discovery cohort comprising metastatic versus benign samples (Table 1) and identified 6 genes (*FAM129A*, *PPA1*, *DPM3*, *SOD2*, *KRT7* and *MME*) for further study (Supplementary Table 1). Of these genes *DPM3*, *SOD2*, *KRT7* and *MME* have all previously been linked to prostate cancer with *DPM3* and *MME* also linked to aggressive disease although no common pathway links these genes

(Ho, et al. 2013; Manos et al. 2001; Voutsadakis, et al. 2012). KRT7 is widely used in cancer diagnosis as a marker for basal cells which are known to be lost in prostate cancer during neoplastic transformation (Brimo and Epstein 2012). Previous studies have suggested that loss of SOD2 expression is an early event in tumour development, providing an environment for increased free radicals and the development of mutations (Hempel, et al. 2011). However, high levels of SOD2 have been reported in metastatic tissue and circulating prostate tumour cells where they are thought to enhance metastases (Giesing, et al. 2012; Hempel et al. 2011). MME protein is reduced in extra-capsular versus organ confined prostate cancer and $MME^{high}/AGR2^{low}$ expression is more closely associated with primary disease whereas $MME^{low}/AGR2^{high}$ is more commonly seen in metastatic tissue (Ho et al. 2013; Kim, et al. 2012). There have also been previous reports of FAM129A overexpression in renal and head and neck cancers where it has been shown to regulate cell stress and p53 degradation to promote cancer cell survival (Adachi, et al. 2004; Cerutti, et al. 2006; Ji, et al. 2012; Matsumoto, et al. 2006; Patel, et al. 2011). We also reported a high FDR for our discovery cohort (0.39) indicating that although we were enriching for differentially expressed genes only 61/100 were likely to be positive. We believe that this is most likely due to the wide range of clinical characteristics in the discovery cohort that had a range of hormone treatments and risk stages. Therefore further examination of any identified genes in well defined cohorts was essential.

Four of these genes (*FAM129A*, *KRT7*, *SOD2* and *MME*), were characterised in circulating mRNA, tissue mRNA and tissue protein levels in a defined hormone cohort and a risk stratified cohort based upon stage, Gleason grade and PSA data (D'Amico et al. 1999) (Figure 1D and Tables 2 and 3). The data generated in all of these cohorts is summarised in Table 5. All of the genes that we identified were down-regulated in prostate cancer which presents challenges in using them as practical biomarkers as technical assay issues can easily give rise to false positive results. The sensitivity of the methodology used for detection could limit the detection of aggressive forms of cancer and loss of signal due to epigenetic changes or mutations cannot always be ruled out.

There was concordance in the expression of *FAM129A*, *KRT7*, *SOD2* and *MME* in circulating mRNA and published metastatic tissue mRNA suggesting that circulating mRNA may accurately reflect

expression that occurs in metastatic tumour tissue either directly or via an immune response (Figures 1C and 1D, Figure 2A) (Taylor et al. 2010). This is consistent with the overexpression of a number of genes identified in circulating tumour cells derived from the tumour tissue (Marin-Aguilera et al. 2015). In our study down-regulation of these four genes could also be confirmed at the protein level in primary tissue biopsies from metastatic men for KRT7 and MME but not SOD2 or FAM129A (Figure 2B). This may reflect the small sample numbers tested or the effects of variable fixation on thin needle core biopsies. These results were supported by the down-regulation of KRT7, SOD2 and MME protein levels in localised tumours which suggests that these 3 genes, and subsequently proteins, are all down-regulated in all tumours compared to benign (Figure 3D and Table 5). Although FAM129A protein was down-regulated in biopsies from metastatic patients this result was not significant ($p=0.17$) and subsequent analysis in localised prostate cancer demonstrated a clear up-regulation in primary tumours (Figures 2B and 3D). For SOD2 the conflicting protein level data between primary biopsy of metastatic disease and localised disease is consistent with other conflicts in metastatic tissue and circulating tumour cells versus primary disease (Giesing et al. 2012; Hempel et al. 2011). There were no significant changes in MME protein levels in both the discovery and defined risk cohorts, as well as in serum protein levels (Figures 1, 2C and 3). This may reflect the greater sensitivity frequently seen in genomic assays (Miura et al. 2005; Papadopoulou et al. 2006; Williams 2010). Despite reports of changes in MME protein in extra-capsular disease we saw no evidence of significant differences between localised and metastatic cohorts. It may be that without the addition of the known prostate biomarker AGR2, MME does not sufficient power alone (Bu, et al. 2011; Kani, et al. 2013).

We hypothesised that some of the differences seen between the metastatic and localised cohorts were as a result of androgen regulation of the identified genes. This was investigated using published ChIP-seq data which showed strong AR binding to the *FAM129A* and *MME* loci in multiple cell lines consistent with previous reports of altered regulation of *FAM129A* in cells treated with the anti-androgen, bicalutamide (Massie et al. 2011; Rothermund, et al. 2005; Wang et al. 2009; Yu et al. 2010). This also suggests that *MME* and *FAM129A* are novel androgen regulated genes and this was

confirmed by demonstrating enrichment of AR binding to the MME, FAM129A and KRT7 promoters following treatment with androgen (Figure 5C). These data suggest that *FAM129A* expression at the mRNA and protein level may initially increase in hormone sensitive tissue before decreasing in patients with hormone refractory disease. This is consistent with an alternatively regulated AR signalling axis and alterations in AR binding sites previously reported in advanced disease (Sharma, et al. 2013; Tsao, et al. 2012). This supports the inclusion of androgen regulated genes in a recently published 17-gene panel to predict aggressive prostate cancer (Klein et al. 2014).

This study has identified a four gene panel of circulating mRNA that can distinguish metastatic from benign disease in a small patient cohort. Although there was frequent concordance between mRNA expression between circulating mRNA and tissue mRNA, there were frequently differences between expression of *FAM129A*, *KRT7*, *SOD2* and *MME* at the mRNA and protein level in many of the cohorts. Androgen regulation of *FAM129A*,*MME* and KRT7 confirms the continuing role of the AR in advanced prostate cancer. This study demonstrates altered AR regulated gene expression during different stages of the prostate cancer disease pathway that can be detected using a minimally invasive blood test. Further work is required to determine if an AR regulated gene panel, either alone or in combination with genes identified in other PAXgene studies, could have clinical utility in predicting aggressive disease or monitoring response to hormone manipulating treatments such as bicalutamide, aberaterone and enzalutamide.

Funding

We acknowledge the support of The University of Cambridge, Cancer Research UK and Hutchison Whampoa Limited in completing this study. We would also like to acknowledge the support of the National Institute for Health Research (NIHR) which funds the Cambridge Bio-medical Research Centre, Cambridge UK and the NCRI (ProMPT) which has funded tissue collections in Cambridge and PCUK that has fund HW and JK for a proportion of the time spent working on this manuscript.

Author contribution statement

BT - expression arrays, qPCR and clinical data analysis, wrote sections of the manuscript, JK – performed haemoglobin reduction experiments, assisted IHC scoring, contributed to data analysis and writing, SM – bioinformatics for haemoglobin reduction and Illumina expression arrays, SV and SD – performed all statistical analysis and contributed to manuscript, LB – performed IHC and assisted in scoring, assisted in the development of the MME ELISA and data analysis, HL – performed the chIP experiments, analysed the data generated and wrote the relevant section of the manuscript, TJ – performed template biopsy versus benign controls, collated and analysed associated clinical data, CM and AL– analysed data from LNCaP expression and chIP-seq studies, provided contributions to the manuscript, MO – expression arrays and advice on haemoglobin reduction experiments, AW – all IHC scoring, producing TMAs, staging and grading all patients, contributed into manuscript preparation, KB – developed and performed MME ELISA on MSD platform, contributed to manuscript writing, AG, JB, MC and SS – consented all patients, collected patient samples, collected and help analyse all clinical data, AL, NS and GS collected samples and analysed clinical data for the HR TMA, DN – helped analyse clinical data and contributed to manuscript, HW – supervised all work, IHC scoring, analysed data, wrote manuscript and contributed some funding.

Acknowledgements

We would like to thank the Genomics Core Facility and Histopathology and In Situ Hybridisation Core Facility, Cancer Research UK Cambridge Institute for their assistance with the work described in this manuscript.

References

Adachi H, Majima S, Kon S, Kobayashi T, Kajino K, Mitani H, Hirayama Y, Shiina H, Igawa M & Hino O 2004 Niban gene is commonly expressed in the renal tumors: a new candidate marker for renal carcinogenesis. *Oncogene* **23** 3495-3500.

Ashida H, Maeda Y & Kinoshita T 2006 DPM1, the catalytic subunit of dolichol-phosphate mannose synthase, is tethered to and stabilized on the endoplasmic reticulum membrane by DPM3. *The Journal of biological chemistry* **281** 896-904.

Augustin H, Mayrhofer K, Pummer K & Mannweiler S 2013 Relationship between prostate cancer gene 3 (PCA3) and characteristics of tumor aggressiveness. *The Prostate* **73** 203-210.

526 Batliwalla FM, Li W, Ritchlin CT, Xiao X, Brenner M, Laragione T, Shao T, Durham R,
527 Kemshetti S, Schwarz E, et al. 2005 Microarray analyses of peripheral blood cells identifies
528 unique gene expression signature in psoriatic arthritis. *Mol Med* **11** 21-29.
529 Bostwick DG, Alexander EE, Singh R, Shan A, Qian J, Santella RM, Oberley LW, Yan T,
530 Zhong W, Jiang X, et al. 2000 Antioxidant enzyme expression and reactive oxygen species
531 damage in prostatic intraepithelial neoplasia and cancer. *Cancer* **89** 123-134.
532 Botchorishvili G, Matikainen MP & Lilja H 2009 Early prostate-specific antigen changes and
533 the diagnosis and prognosis of prostate cancer. *Current opinion in urology* **19** 221-226.
534 Brimo F & Epstein JI 2012 Immunohistochemical pitfalls in prostate pathology. *Human*
535 *pathology* **43** 313-324.
536 Bu H, Bormann S, Schafer G, Horninger W, Massoner P, Neeb A, Lakshmanan VK,
537 Maddalo D, Nestl A, Sultmann H, et al. 2011 The anterior gradient 2 (AGR2) gene is
538 overexpressed in prostate cancer and may be useful as a urine sediment marker for prostate
539 cancer detection. *The Prostate* **71** 575-587.
540 Cairns J, Dunning MJ, Ritchie ME, Russell R, and Lynch AG. 2008 BASH: a tool for
541 managing BeadArray spatial artefacts. . *Bioinformatics* **24 (24)** 2921– 2922.
542 Catalona WJ, Hudson MA, Scardino PT, Richie JP, Ahmann FR, Flanigan RC, deKernion
543 JB, Ratliff TL, Kavoussi LR, Dalkin BL, et al. 1994 Selection of optimal prostate specific
544 antigen cutoffs for early detection of prostate cancer: receiver operating characteristic curves.
545 *J Urol* **152** 2037-2042.
546 Cerutti JM, Latini FR, Nakabashi C, Delcelo R, Andrade VP, Amadei MJ, Maciel RM, Hojaij
547 FC, Hollis D, Shoemaker J, et al. 2006 Diagnosis of suspicious thyroid nodules using four
548 protein biomarkers. *Clinical cancer research : an official journal of the American*
549 *Association for Cancer Research* **12** 3311-3318.
550 Claessens F, Helsen C, Prekovic S, Van den Broeck T, Spans L, Van Poppel H & Joniau S
551 2014 Emerging mechanisms of enzalutamide resistance in prostate cancer. *Nature reviews.*
552 *Urology*.
553 D'Amico AV, Whittington R, Malkowicz SB, Fondurulia J, Chen MH, Kaplan I, Beard CJ,
554 Tomaszewski JE, Renshaw AA, Wein A, et al. 1999 Pretreatment nomogram for prostate-
555 specific antigen recurrence after radical prostatectomy or external-beam radiation therapy for
556 clinically localized prostate cancer. *J Clin Oncol* **17** 168-172.
557 Danila DC, Anand A, Schultz N, Heller G, Wan M, Sung CC, Dai C, Khanin R, Fleisher M,
558 Lilja H, et al. 2013 Analytic and Clinical Validation of a Prostate Cancer-Enhanced
559 Messenger RNA Detection Assay in Whole Blood as a Prognostic Biomarker for Survival.
560 *European urology*.
561 de Kok JB, Verhaegh GW, Roelofs RW, Hessels D, Kiemeney LA, Aalders TW, Swinkels
562 DW & Schalken JA 2002 DD3(PCA3), a very sensitive and specific marker to detect prostate
563 tumors. *Cancer Res* **62** 2695-2698.
564 Dunning M, Smith ML, Ritchie ME, and Tavare S 2007 beadarray: R classes and methods for
565 Illumina bead-based data. . *Bioinformatics* **23(16)** 2183–2184.
566 Egger M, Zwahlen M & Cerny T 2013 [Cancer screening - principles, evaluation and
567 implementation]. *Therapeutische Umschau. Revue therapeutique* **70** 195-203.
568 Fan H & Hegde PS 2005 The transcriptome in blood: challenges and solutions for robust
569 expression profiling. *Curr Mol Med* **5** 3-10.
570 Gentleman R, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge
571 Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M,
572 Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JYH , and Zhang J. 2004
573 Bioconductor: Open software development for computational biology and bioinformatics.
574 *Genome Biology* **5** R80.

Giesing M, Driesel G, Molitor D & Suchy B 2012 Molecular phenotyping of circulating tumour cells in patients with prostate cancer: prediction of distant metastases. *BJU international* **110** E1202-1211.
Glass C & Fuchs E 1988 Isolation, sequence, and differential expression of a human K7 gene in simple epithelial cells. *The Journal of cell biology* **107** 1337-1350.
Glinsky GV, Glinskii AB, Stephenson AJ, Hoffman RM & Gerald WL 2004 Gene expression profiling predicts clinical outcome of prostate cancer. *The Journal of clinical investigation* **113** 913-923.
Haese A, de la Taille A, van Poppel H, Marberger M, Stenzl A, Mulders PF, Huland H, Abbou CC, Remzi M, Tinzl M, et al. 2008 Clinical utility of the PCA3 urine assay in European men scheduled for repeat biopsy. *European urology* **54** 1081-1088.
Hempel N, Carrico PM & Melendez JA 2011 Manganese superoxide dismutase (Sod2) and redox-control of signaling events that drive metastasis. *Anti-cancer agents in medicinal chemistry* **11** 191-201.
Hessels D & Schalken JA 2009 The use of PCA3 in the diagnosis of prostate cancer. *Nature reviews. Urology* **6** 255-261.
Ho ME, Quek SI, True LD, Morrissey C, Corey E, Vessella RL, Dumpit R, Nelson PS, Maresh EL, Mah V, et al. 2013 Prostate cancer cell phenotypes based on AGR2 and CD10 expression. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc* **26** 849-859.
Hothorn T, Hornik K and Zeileis A 2006 Recursive Partitioning: A Conditional Inference Framework. *Journal of Computational and Graphical Statistics* **15(3)** 651--674.
Ji H, Ding Z, Hawke D, Xing D, Jiang BH, Mills GB & Lu Z 2012 AKT-dependent phosphorylation of Niban regulates nucleophosmin- and MDM2-mediated p53 stability and cell apoptosis. *EMBO reports* **13** 554-560.
Kani K, Malihi PD, Jiang Y, Wang H, Wang Y, Ruderman DL, Agus DB, Mallick P & Gross ME 2013 Anterior gradient 2 (AGR2): blood-based biomarker elevated in metastatic prostate cancer associated with the neuroendocrine phenotype. *The Prostate* **73** 306-315.
Kattan MW, Eastham JA, Stapleton AM, Wheeler TM & Scardino PT 1998 A preoperative nomogram for disease recurrence following radical prostatectomy for prostate cancer. *J Natl Cancer Inst* **90** 766-771.
Kim Y, Ignatchenko V, Yao CQ, Kalatskaya I, Nyalwidhe JO, Lance RS, Gramolini AO, Troyer DA, Stein LD, Boutros PC, et al. 2012 Identification of differentially expressed proteins in direct expressed prostatic secretions of men with organ-confined versus extracapsular prostate cancer. *Molecular & cellular proteomics : MCP* **11** 1870-1884.
Kitchen RR, Sabine VS, Simen AA, Dixon JM, Bartlett JM & Sims AH 2011 Relative impact of key sources of systematic noise in Affymetrix and Illumina gene-expression microarray experiments. *BMC genomics* **12** 589.
Klein EA, Cooperberg MR, Magi-Galluzzi C, Simko JP, Falzarano SM, Maddala T, Chan JM, Li J, Cowan JE, Tsiatis AC, et al. 2014 A 17-gene Assay to Predict Prostate Cancer Aggressiveness in the Context of Gleason Grade Heterogeneity, Tumor Multifocality, and Biopsy Undersampling. *European urology*.
Leon SA, Shapiro B, Sklaroff DM & Yaros MJ 1977 Free DNA in the serum of cancer patients and the effect of therapy. *Cancer Res* **37** 646-650.
Lewis DA, Stashenko GJ, Akay OM, Price LI, Owzar K, Ginsburg GS, Chi JT & Ortel TL 2011 Whole blood gene expression analyses in patients with single versus recurrent venous thromboembolism. *Thromb Res* **128** 536-540.
Li D, Butt A, Clarke S & Swaminathana R 2004 Real-time quantitative PCR measurement of thyroglobulin mRNA in peripheral blood of thyroid cancer patients and healthy subjects. *Ann N Y Acad Sci* **1022** 147-151.

Liong ML, Lim CR, Yang H, Chao S, Bong CW, Leong WS, Das PK, Loh CS, Lau BE, Yu CG, et al. 2012 Blood-based biomarkers of aggressive prostate cancer. *PloS one* **7** e45802.
Manos EJ, Kim ML, Kassis J, Chang PY, Wells A & Jones DA 2001 Dolichol-phosphate-mannose-3 (DPM3)prostin-1 is a novel phospholipase C-gamma regulated gene negatively associated with prostate tumor invasion. *Oncogene* **20** 2781-2790.
Marin-Aguilera M, Reig O, Lozano JJ, Jimenez N, Garcia-Recio S, Erill N, Gaba L, Tagliapietra A, Ortega V, Carrera G, et al. 2015 Molecular profiling of peripheral blood is associated with circulating tumor cells content and poor survival in metastatic castration-resistant prostate cancer. *Oncotarget* **6** 10604-10616.
Massie CE, Lynch A, Ramos-Montoya A, Boren J, Stark R, Fazli L, Warren A, Scott H, Madhu B, Sharma N, et al. 2011 The androgen receptor fuels prostate cancer by regulating central metabolism and biosynthesis. *The EMBO journal* **30** 2719-2733.
Massie CE & Mills IG 2009 Chromatin immunoprecipitation (ChIP) methodology and readouts. *Methods in molecular biology* **505** 123-137.
Mateo J, Smith A, Ong M & de Bono JS 2014 Novel drugs targeting the androgen receptor pathway in prostate cancer. *Cancer metastasis reviews*.
Matsumoto F, Fujii H, Abe M, Kajino K, Kobayashi T, Matsumoto T, Ikeda K & Hino O 2006 A novel tumor marker, Niban, is expressed in subsets of thyroid tumors and Hashimoto's thyroiditis. *Human pathology* **37** 1592-1600.
Miura N, Maeda Y, Kanbe T, Yazama H, Takeda Y, Sato R, Tsukamoto T, Sato E, Marumoto A, Harada T, et al. 2005 Serum human telomerase reverse transcriptase messenger RNA as a novel tumor marker for hepatocellular carcinoma. *Clinical cancer research : an official journal of the American Association for Cancer Research* **11** 3205-3209.
Olmos D, Brewer D, Clark J, Danila DC, Parker C, Attard G, Fleisher M, Reid AH, Castro E, Sandhu SK, et al. 2012 Prognostic value of blood mRNA expression signatures in castration-resistant prostate cancer: a prospective, two-stage study. *The lancet oncology*.
Papadopoulou E, Davilas E, Sotiriou V, Georgakopoulos E, Georgakopoulou S, Koliopanos A, Aggelakis F, Dardoufas K, Agnanti NJ, Karydas I, et al. 2006 Cell-free DNA and RNA in plasma as a new molecular marker for prostate and breast cancer. *Ann N Y Acad Sci* **1075** 235-243.
Partin AW, Yoo J, Carter HB, Pearson JD, Chan DW, Epstein JI & Walsh PC 1993 The use of prostate specific antigen, clinical stage and Gleason score to predict pathological stage in men with localized prostate cancer. *J Urol* **150** 110-114.
Patel MR, Stadler ME, Deal AM, Kim HS, Shores CG & Zanation AM 2011 STT3A, C1orf24, TFF3: putative markers for characterization of follicular thyroid neoplasms from fine-needle aspirates. *The Laryngoscope* **121** 983-989.
Quek SI, Ho ME, Loprieno MA, Ellis WJ, Elliott N & Liu AY 2012 A multiplex assay to measure RNA transcripts of prostate cancer in urine. *PloS one* **7** e45656.
Rainen L, Oelmueller U, Jurgensen S, Wyrich R, Ballas C, Schram J, Herdman C, Bankaitis-Davis D, Nicholls N, Trollinger D, et al. 2002 Stabilization of mRNA expression in whole blood samples. *Clin Chem* **48** 1883-1890.
Ramos-Montoya A, Lamb A, Russell R, Carroll T, Jurmeister S, Galeano-Dalmau N, Massie C, Boren J, Bon H, Theodorou V, et al. 2014 Hes6 drives a critical AR transcriptional program to induce castration resistant
prostate cancer through activation of an E2F1-mediated cell cycle network. *EMBO Molecular Medicine* **In press.**
Ross RW, Galsky MD, Scher HI, Magidson J, Wassmann K, Lee GS, Katz L, Subudhi SK, Anand A, Fleisher M, et al. 2012 A whole-blood RNA transcript-based prognostic model in men with castration-resistant prostate cancer: a prospective study. *The lancet oncology*.

674 Rothermund CA, Gopalakrishnan VK, Eudy JD & Vishwanatha JK 2005 Casodex treatment
675 induces hypoxia-related gene expression in the LNCaP prostate cancer progression model.
676 *BMC urology* **5** 5.
677 Schwarzenbach H, Hoon DS & Pantel K 2011 Cell-free nucleic acids as biomarkers in cancer
678 patients. *Nat Rev Cancer* **11** 426-437.
679 Shariat SF & Roehrborn CG 2008 Using biopsy to detect prostate cancer. *Rev Urol* **10** 262-
680 280.
681 Sharma NL, Massie CE, Ramos-Montoya A, Zecchini V, Scott HE, Lamb AD, MacArthur S,
682 Stark R, Warren AY, Mills IG, et al. 2013 The androgen receptor induces a distinct
683 transcriptional program in castration-resistant prostate cancer in man. *Cancer cell* **23** 35-47.
684 Smyth G 2005 Limma: linear models for microarray data. In *Bioinformatics and*
685 *Computational Biology Solutions using R and Bioconductor*, pp 397-420. Ed CV Gentleman
686 R, Dudoit S, and Huber W. New York: Springer.
687 Statistics OfN 2013 Prostate cancer: the most common cancer among men in England - 2010
688 London.
689 Taylor BS, Schultz N, Hieronymus H, Gopalan A, Xiao Y, Carver BS, Arora VK, Kaushik P,
690 Cerami E, Reva B, et al. 2010 Integrative genomic profiling of human prostate cancer.
691 *Cancer cell* **18** 11-22.
692 Thompson IM, Ankerst DP, Chi C, Lucia MS, Goodman PJ, Crowley JJ, Parnes HL &
693 Coltman CA, Jr. 2005 Operating characteristics of prostate-specific antigen in men with an
694 initial PSA level of 3.0 ng/ml or lower. *JAMA* **294** 66-70.
695 Tsao CK, Galsky MD, Small AC, Yee T & Oh WK 2012 Targeting the androgen receptor
696 signalling axis in castration-resistant prostate cancer (CRPC). *BJU international* **110** 1580-
697 1588.
698 Vartanian K, Slottke R, Johnstone T, Casale A, Planck SR, Choi D, Smith JR, Rosenbaum JT
699 & Harrington CA 2009 Gene expression profiling of whole blood: comparison of target
700 preparation methods for accurate and reproducible microarray analysis. *BMC genomics* **10** 2.
701 Voutsadakis IA, Vlachostergios PJ, Daliani DD, Karasavvidou F, Kakkas G, Moutzouris G,
702 Melekos MD & Papandreou CN 2012 CD10 is inversely associated with nuclear factor-kappa
703 B and predicts biochemical recurrence after radical prostatectomy. *Urologia internationalis*
704 **88** 158-164.
705 Wang Q, Li W, Zhang Y, Yuan X, Xu K, Yu J, Chen Z, Beroukhim R, Wang H, Lupien M, et
706 al. 2009 Androgen receptor regulates a distinct transcription program in androgen-
707 independent prostate cancer. *Cell* **138** 245-256.
708 Whitaker HC, Kote-Jarai Z, Ross-Adams H, Warren AY, Burge J, George A, Bancroft E,
709 Jhavar S, Leongamornlert D, Tymrakiewicz M, et al. 2010 The rs10993994 risk allele for
710 prostate cancer results in clinically relevant changes in microseminoprotein-beta expression
711 in tissue and urine. *PloS one* **5** e13363.
712 Whitaker HC, Patel D, Howat WJ, Warren AY, Kay JD, Sangan T, Marioni JC, Mitchell J,
713 Aldridge S, Luxton HJ, et al. 2013 Peroxiredoxin-3 is overexpressed in prostate cancer and
714 promotes cancer cell survival by protecting cells from oxidative stress. *British journal of*
715 *cancer* **109** 983-993.
716 Williams MA 2010 Stabilizing the code-methods to preserve RNA prove their worth.
717 *Biomark Insights* **5** 139-143.
718 Wispe JR, Clark JC, Burhans MS, Kropp KE, Korfhagen TR & Whitsett JA 1989 Synthesis
719 and processing of the precursor for human mangano-superoxide dismutase. *Biochimica et*
720 *biophysica acta* **994** 30-36.
721 Yang B, Xu Q, Wu F, Liu F, Ye X, Liu G, Shao Z, Meng X, Mougin B & Wu J 2011 Using
722 peripheral blood mRNA signature to distinguish between breast cancer and benign breast
723 disease in non-conclusive mammography patients. *Cancer Biol Ther* **10** 1235-1239.

Yu J, Mani RS, Cao Q, Brenner CJ, Cao X, Wang X, Wu L, Li J, Hu M, Gong Y, et al. 2010 An integrated network of androgen receptor, polycomb, and TMPRSS2-ERG gene fusions in prostate cancer progression. *Cancer cell* **17** 443-454.
Zheng R, Horiguchi A, Iida K, Lee J, Shen R, Goodman OB, Jr. & Nanus DM 2010 Neutral endopeptidase is a myristoylated protein. *Molecular and cellular biochemistry* **335** 173-180.

Figure legends

Figure 1: Identification of differentially expressed genes in circulating mRNA. (A) A schematic overview of the study design giving numbers of patients and genes included at each stage (n). (B) A heat map of clustered results from the HT12 expression arrays for the 6 identified genes (*SOD2*, *FAM129A*, *MME*, *KRT7*, *PPA1* and *DPM3*). Multiple probes for these genes that appeared in the top 1000 most DE probes are shown. Red indicates high expression while blue indicates low expression. (C) Boxplots showing the expression array data for each probe shown in (B), where multiple probes exist for a gene all probes are shown. *Limma* p-values for diffential gene expression are shown. (D) qPCR validation of expression array results. Ct values were calculated for all conditions and the expression of target genes was normalised against the expression of *RPLP2* housekeeping gene using the δδCt method. All primer sequences are given in Supplementary Table 5. All p-values were calculated using a Mann Whitney 2-tailed t-test. The control group in all Figures were patients with raised PSA, negative biopsy.

Figure 2: Expression of FAM129A, KRT7, SOD2 and MME in tissue and serum from metastatic patients. (A) Box and whisker plots were generated from previously published data on metastatic and control cohorts (Taylor et al. 2010). P-values were calculated using a Wilcoxon rank-sum test. (B) Immunohistochemistry (IHC) for the 4 genes on biopsy tissue taken from the patients assayed in Figure 1. All IHC was performed on the Bondmax Autostainer using the antibodies FAM129A (1:750), KRT7 (2.8g/L), SOD2 (1:1500) and MME (1:50). Staining is shown in brown with nuclei shown in blue and was classified into the following categories: none, weak, moderate and high, based on intensity. (C) An ELISA developed on the MesoScale Discovery platform was used to determine the amount of MME protein in the serum from patients previously tested in the discovery cohort. Plates were coated with

goat anti-MME (1.44µg/ml) and MME protein detected with biotinylated goat anti-MME (1:100) before visualising using streptavidin sulphoTAG (1:1000) and measured using a Sector 6000 plate reader.

Figure 3: Expression of FAM129A, KRT7, SOD2 and MME in different risk cohorts. Patients were defined as low, intermediate or high risk based on the following clinical criteria; low risk – PSA ≤10ng/ml, Gleason 6, ≥T2, intermediate risk – PSA ≥10ng/ml but ≤20ng/ml, Gleason 7, ≥T2 or high risk – PSA ≥20ng/ml, Gleason ≥8, ≥T3. The clinical characteristics of the risk cohort are given in Supplementary Table 3. (A)The circulating mRNA for the 4 genes was determined in 3 risk cohorts by qPCR. Ct values were calculated for all conditions and the expression of target genes was normalised against the expression of *RPLP2* housekeeping gene using the δδCt method. Kruskal Wallis tests for each gene testing the likelihood of a result falling into any group by chance gave the following results; *FAM129A* – p=0.34, *KRT7* – p=0.0017, *SOD2* – p=0.79 and *MME* p=0.033. All p-values were calculated using a Mann Whitney 2-tailed t-test. (B) Recursive partitioning was performed using Ct values from the qPCR validation to predict cut-offs for each group within the data. Only genes showing significant results are shown. Ct cut-off values and 95% confidence intervals are indicated. High – high risk, inter – intermediate risk, low – low risk. (C) Expression of *FAM129A*, *KRT7*, *SOD2* and *MME* was determined in localised prostate using published expression data (Taylor et al. 2010) and p-values determined using a Wilcoxon rank-sum test (D) Alterations in the protein levels of FAM129A, KRT7, SOD2 and MME in localised disease was determined by IHC using an in-house TMA previously described where G3, G4 and G5 refer to Gleason grades (Whitaker et al. 2010; Whitaker et al. 2013). All IHC was performed on the Bondmax Autostainer using conditions described before. Staining is shown in brown with nuclei shown in blue and was classified into the following categories: none, weak, moderate and high, based on intensity. P-values were calculated using a Kruskal Wallis test.

Figure 4: Hormone regulation of FAM129A, KRT7, SOD2 and MME in patient cohorts. (A) The circulating mRNA for the 4 genes was determined in 3 hormone cohorts by qPCR; patients yet to receive treatment (hormone naive), receiving hormone treatment and continuing to respond (hormonal therapy) and patients who have become hormone refractory (hormone relapsed). Ct values were

781 calculated for all conditions and the expression of target genes was normalised against the expression
782 of *RPLP2* housekeeping gene using the δδCt method. Kruskal Wallis tests for each gene testing the
783 likelihood of a result falling into any group by chance gave the following results; FAM129 –
784 p=0.0085, *KRT7* – p=0.038, *SOD2* – p=0.014 and *MME* p=0.0079. All p-values shown were
785 calculated using a Mann Whitney 2-tailed t-test between two groups. (B) Recursive partitioning was
786 performed using Ct values from the qPCR validation to predict cut-offs for each group within the
787 data. Only genes showing significant results are shown. Ct cut-off values and 95% confidence
788 intervals are indicated. HN – never received hormone treatment (hormone naïve), HT – receiving
789 hormone treatment and responding (hormonal therapy) and HR – receiving hormone treatment and
790 showing biochemical or clinical relapse (hormone relapsed). (C) Alterations in the protein levels of
791 FAM129A, KRT7, SOD2 and MME with hormone status was determined by IHC using a hormone
792 refractory TMA described in Supplementary Table 7. All IHC was performed on the Bondmax
793 Autostainer using conditions described before. Staining is shown in brown with nuclei shown in blue
794 and was classified into the following categories: none, weak, moderate and high, based on intensity.
795 P-values across all groups were calculated using a Kruskal-Wallis test and pairwise comparisons
796 using a Mann-Whitney two-tailed t-test.
797
798 *Figure 5: Androgen regulation of the 4 gene panel.* (A) Androgen regulation of the 4 genes was
799 determined in the LNCaP cell line by treating with androgens and taking samples for expression
800 analysis over the following 24 hours (Massie et al. 2011). Filled circles represent androgen treatment,
801 open circles represent vehicle controls. Data for all good probes is shown for *FAM129A*. Data for
802 other genes is shown in Supplementary Figure 6. ACF = autocorrelation (function). A measure greater
803 than zero means that consecutive time points are nearer each other than time points chosen at random.
804 (B) UCSC genome browser view of the *FAM129A* and *MME* loci showing AR binding profiles in
805 prostate cancer cells from three independent studies (Massie et al. 2011; Wang et al. 2009; Yu et al.
806 2010). Coloured blocks indicate AR peaks identified in each of the three chIP studies, red peaks show
807 the AR chIP-seq profile for two cell lines (Massie et al. 2011). Below gene annotations are ENCODE
808 tracks indicating promotor, enhancer, DNAse hypersensitivity and transcription factor binding

809 profiles. Arrows indicate promoters and direction of transcription. Data for *KRT7* and *SOD2* are
810 shown in Supplementary Figure 7. (C) chIP for the AR in the LNCaP cell line following starvation for
811 48 hours in charcoal stripped media and treatment with 10^{-8}M R1881 for 1 hour (black bars). Data is
812 shown as relative to vehicle treated cells (grey bars). P-values are calculated using a Wilcoxon rank-
813 sum test
814
815 *Table 1:* Clinical data for the Discovery cohort and template biopsy confirmed benign cohort.
816 N=number of patients with percentage given in in brackets afterwards. *Age and PSA data collected
817 at time of diagnosis.
818
819 *Table 2:* Clinical data for the defined hormone status cohort. N=number of patients with percentage
820 given in in brackets afterwards. *Age and PSA data collected at time of diagnosis.
821
822 *Table 3:* Clinical data for the risk stratified. N=number of patients with percentage given in in
823 brackets afterwards. *Age and PSA data collected at time of diagnosis. The median proportion of
824 positive cores is also given for each group with the range shown in brackets.
825
826 *Table 4:* Clinical data for the hormone relapsed (HR) TMA generated in Cambridge. N=number of
827 patients with percentage given in in brackets afterwards.
828
829
830 *Table 5:* An overview of the data generated at the mRNA and protein level in the three different
831 cohorts used in this study. Intermediate (inter) versus high risk (high) is shown for the risk cohort. A
832 statistically significant decrease or increase is shown by ↓ (blue shading) and ↑ (red shading)
833 respectively. Data that was not significant is shown as ↔ (yellow shading).
834
835 *Supplementary Figure 1:* Quality control data of signal versus noise from the expression arrays of
836 globin-reduced and non-reduced samples. The X-axis represents P95 (signal) values and the Y-Axis

represents P05 (noise) values. The red line represents a signal to noise ratio of 10 which is recommended by the array manufacturers (B) and (C) Expression versus variability of a set of classic housekeeping genes is shown in the globin-reduced and non-reduced expression data. The horizontal lines represent the median values of the respective distributions. RPLP2 was selected as the control gene.

Supplementary Figure 2: siRNA targeting *FAM129A*, *SOD2* and *MME or a non-targeting control (NT)* was used to transiently knock-down each gene in LNCaP cells for 72 hours. Half of the cells were harvested for qPCR validation of knock-down (A) while the remainder were loosely pelleted, formalin fixed for 16 hours and made into a paraffin embedded cell pellet. Sections of this pellet were stained by IHC for the respectively gene (B).

Supplementary Figure 3: Haematological parameters including haemoglobin, white cell count (WCC) and platelets for patients included in the Discovery Cohort described in Supplementary Table 1. The benign cohort represents those patients with raised PSA and negative TRUS biopsy.

Supplementary Figure 4: A heatmap showing the 100 differentially expressed genes in the discovery cohort as determined using the non-globin reduced samples. Differential expression analysis was carried out using the *limma* package and genes selected for further analysis by ranking according to Benjamini-Hochberg calculated, unadjusted, p-values and a cut off of 0.01. Blue represents down-regulated genes and red overexpressed genes.

Supplementary Figure 5: qPCR of circulating mRNA for *FAM129A*, *KRT7*, *SOD2* and *MME* from two benign cohorts; raised PSA negative template biopsy and raised PSA negative TRUS biopsy. Ct values were calculated for all conditions and the expression of target genes was normalised against the expression of *RPLP2* housekeeping gene using the δδCt method. P-values were calculated using a Mann Whitney two-tailed t-test.

Supplementary Figure 6: Kaplan Meier curves for FAM129A, *KRT7*, *SOD2* and *MME* in the Taylor et al. (Taylor et al. 2010) (A) or Glinsky et al. (Glinsky et al. 2004) (B) datasets generated using recursive partitioning. Red indicates high expression and blue indicates low expression. P-values were calculated using a Jonckheere-Terpstra test.

Supplementary Figure 7: Androgen regulation of the 4 genes was determined in the LNCaP cell line by treating with androgens and taking samples for expression analysis over the following 24 hours (Massie et al. 2011). Filled circles represent androgen treatment, open circles represent vehicle controls. Data for all probes is shown. UCSC genome browser view of the *FAM129A* and *MME* loci showing AR binding profiles in prostate cancer cells from three independent studies (Massie et al. 2011; Wang et al. 2009; Yu et al. 2010). Coloured blocks indicate AR peaks identified in each of the three chIP studies, red peaks show the AR chIP-seq profile for two cell lines (Massie et al. 2011). Below gene annotations are ENCODE tracks indicating promoter, enhancer, DNAse hypersensitivity and transcription factor binding profiles. Arrows indicate promoters and direction of transcription.

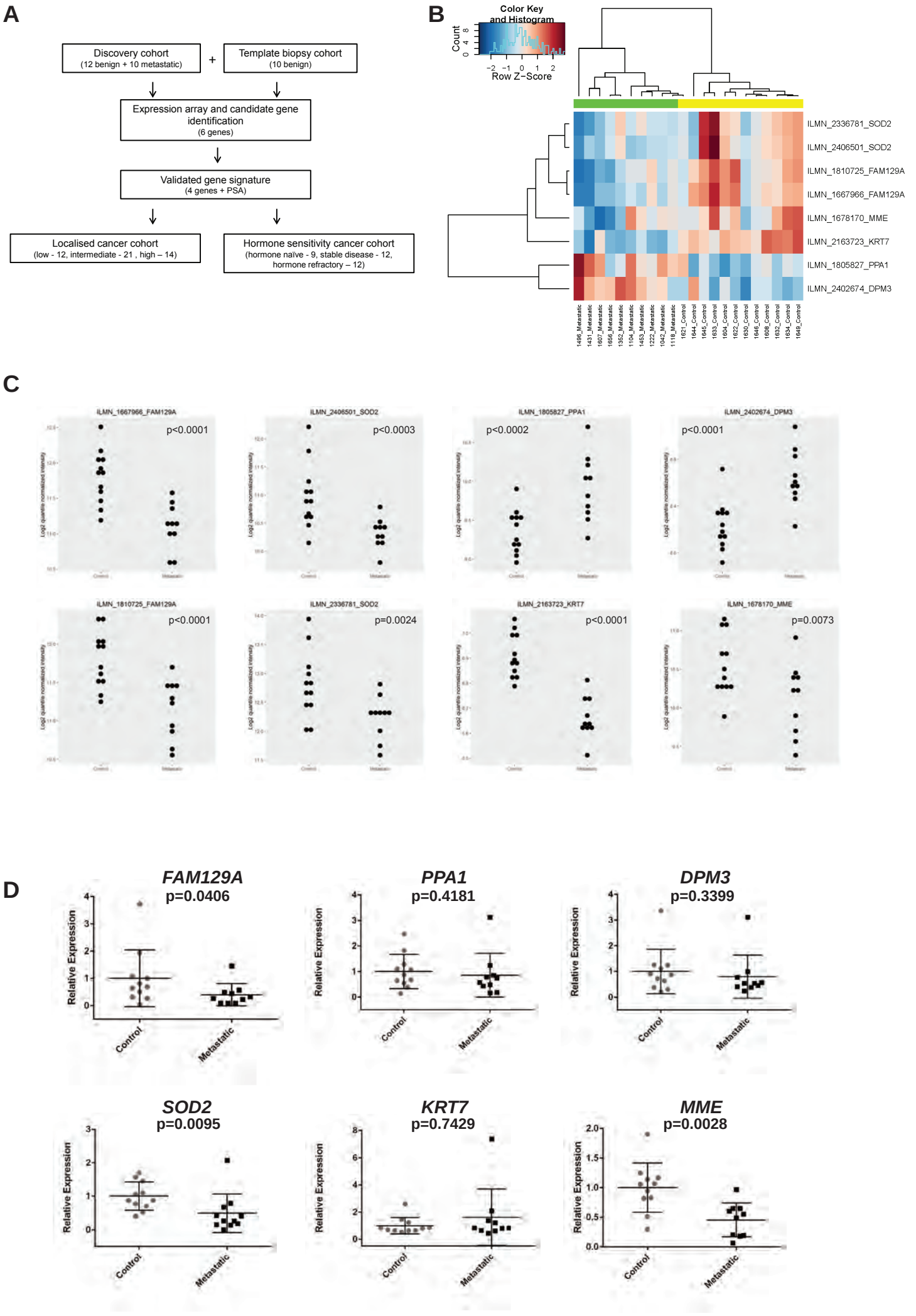
Supplementary Table 1: Overview of genes identified in HT12 arrays. A brief overview of current knowledge is given under 'function' and any known links to any cancer and specifically prostate cancer.

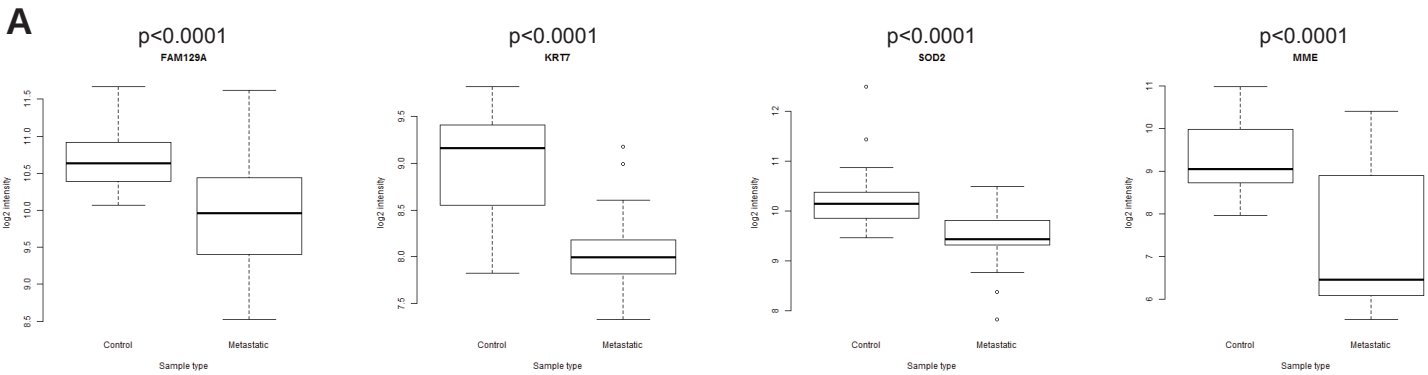
Supplementary Table 2: Genes to be validated in additional datasets were identified by expression array analysis of non-haemoglobin reduced samples. Differential expression analysis was carried out using the *limma* package and genes selected for further analysis by ranking according to Benjamini-Hochberg calculated, unadjusted, p-values and a cut off of 0.01. Genes with a significant association with poor outcome in either Taylor or Glinsky dataset are highlighted in red. Preference was given to genes with good antibody availability. Antibodies were ranked by HW according to published staining examples in prostate cancer; 5* - strong, crisp staining, no background, demonstrates clear ability to discriminate between benign and tumour cells, 4* - strong staining, low non-specific binding or limited ability to discriminate between benign and tumour cells, 3* - moderate staining with

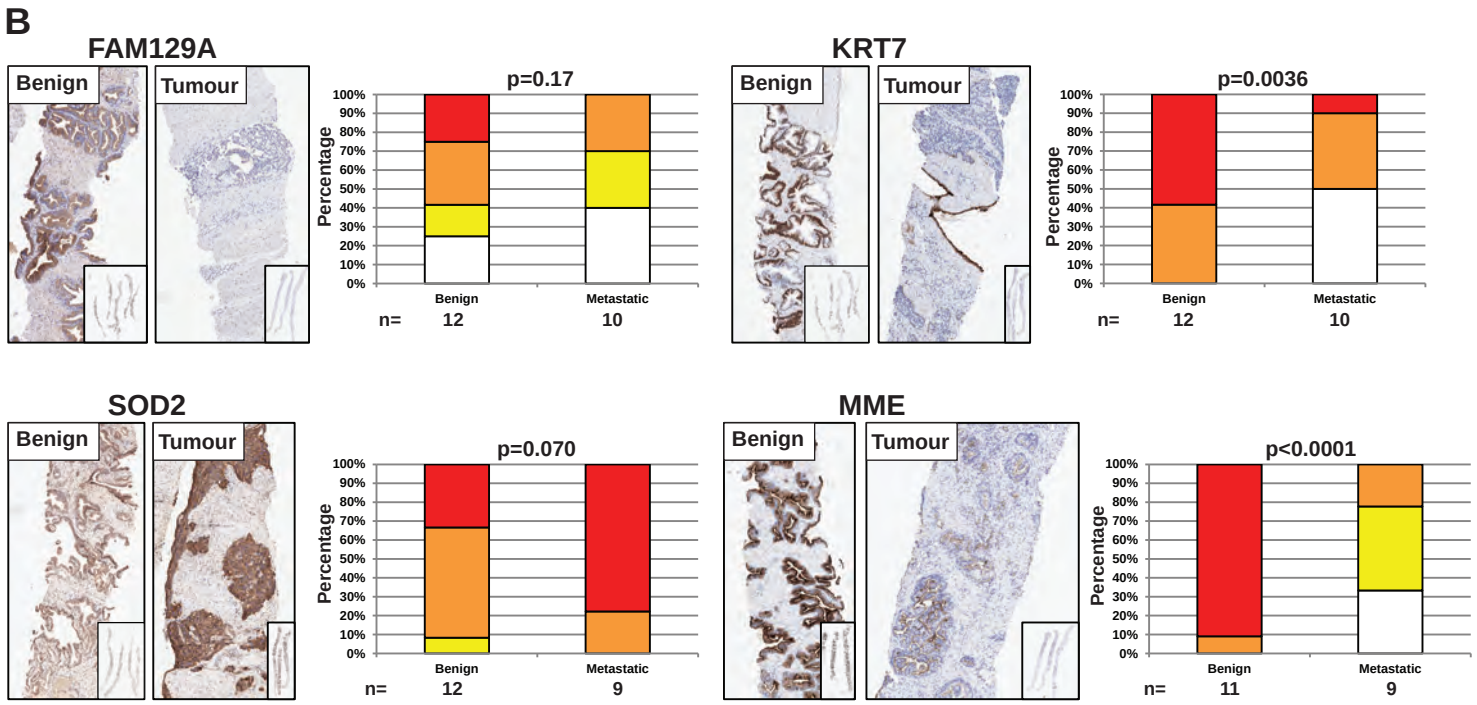
moderate non-specific staining, 2* - weak or moderate staining with high levels of non-specific staining, 1* - no or weak staining seen, N/a - no antibody staining pictures available for assessment. Genes taken forward in this study are highlighted in yellow.

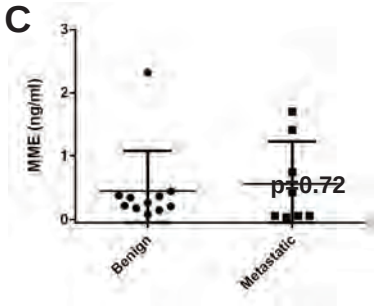
Supplementary Table 3: Primers used for qPCR validation of genes identified in the expression array. All sequences are shown 5'to 3'.

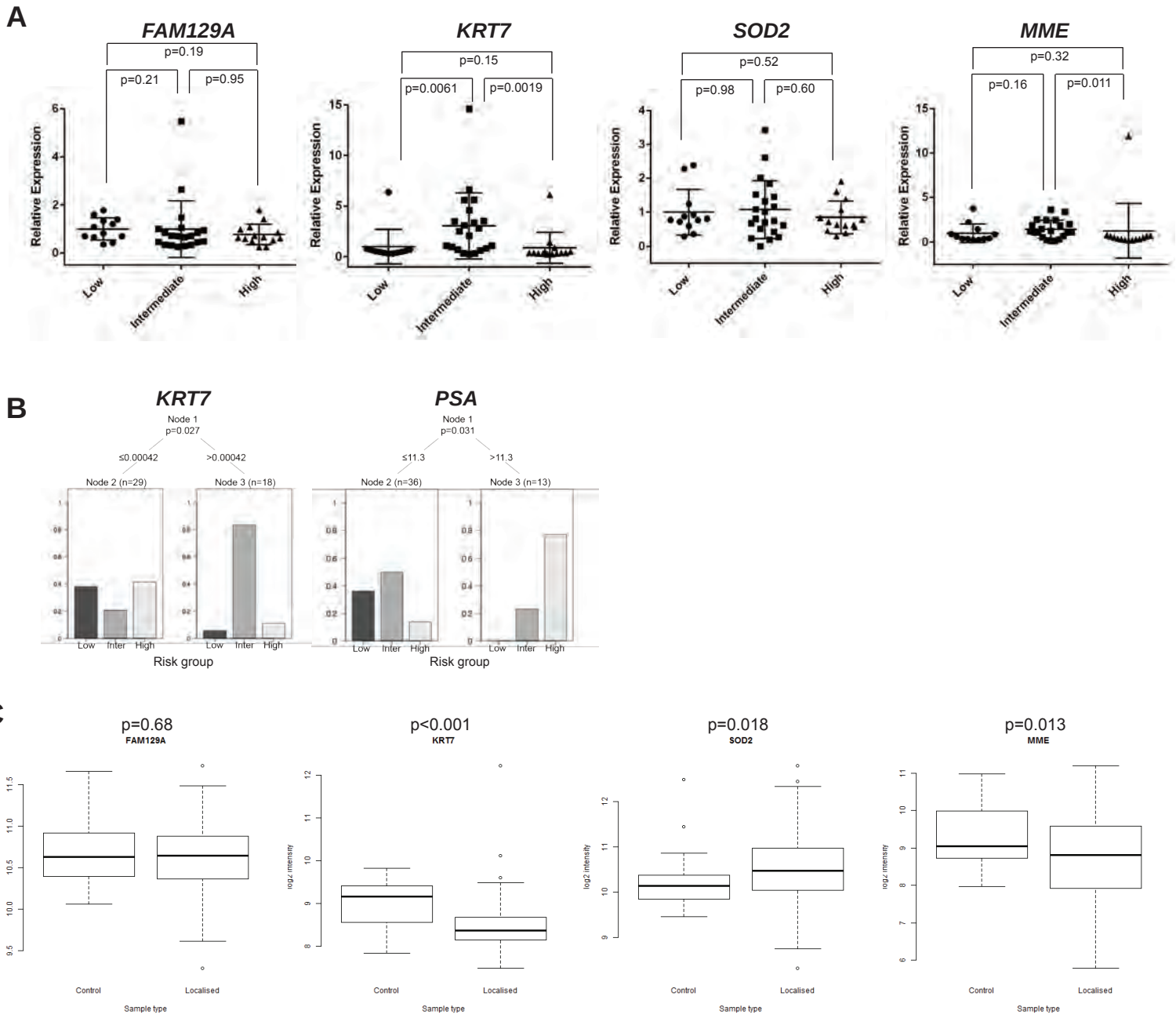
Supplementary Table 4: Primers used for chIP experiment. All sequences are shown 5'to 3'.

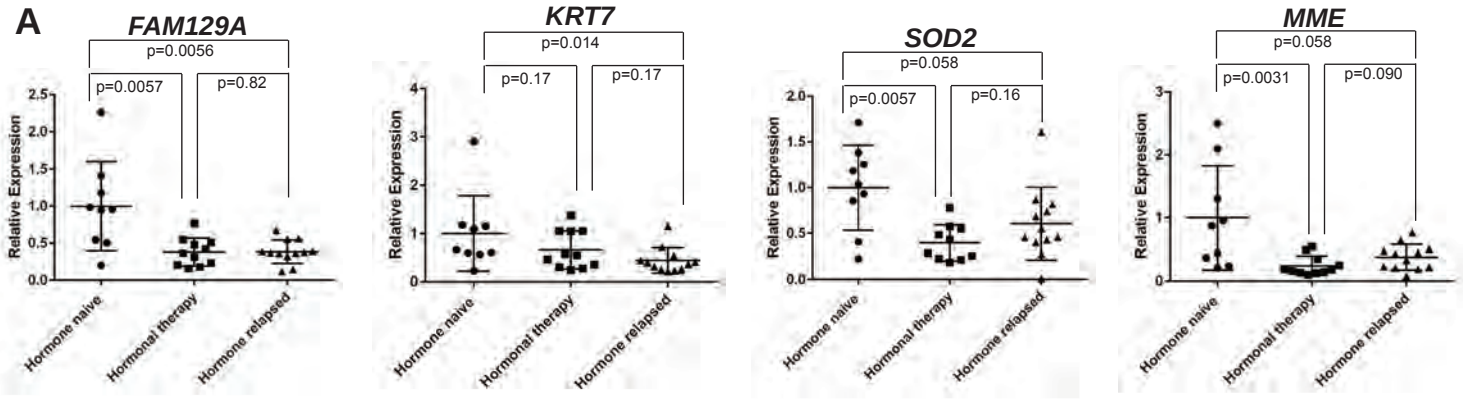

A
Discovery cohort
(12 benign + 10 metastatic)
+
Template biopsy cohort
(10 benign)
Expression array and candidate gene identification
(6 genes)
Validated gene signature
(4 genes + PSA)
Localised cancer cohort
(low - 12, intermediate - 21 , high – 14)
Hormone sensitivity cancer cohort
(hormone naïve - 9, stable disease - 12, hormone refractory – 12)
B
Color Key and Histogram
Count
Row Z-Score
ILMN_2336781_SOD2
ILMN_2406501_SOD2
ILMN_1810725_FAM129A
ILMN_1667966_FAM129A
ILMN_1678170_MME
ILMN_2163723_KRT7
ILMN_1805827_PPA1
ILMN_2402674_DPM3
C
ILMN_1667966_FAM129A p<0.0001
ILMN_2406501_SOD2 p<0.0003
ILMN_1805827_PPA1 p<0.0002
ILMN_2402674_DPM3 p<0.0001
ILMN_1810725_FAM129A p<0.0001
ILMN_2336781_SOD2 p=0.0024
ILMN_2163723_KRT7 p<0.0001
ILMN_1678170_MME p=0.0073
D
FAM129A p=0.0406
PPA1 p=0.4181
DPM3 p=0.3399
SOD2 p=0.0095
KRT7 p=0.7429
MME p=0.0028
Relative Expression
Control
Metastatic

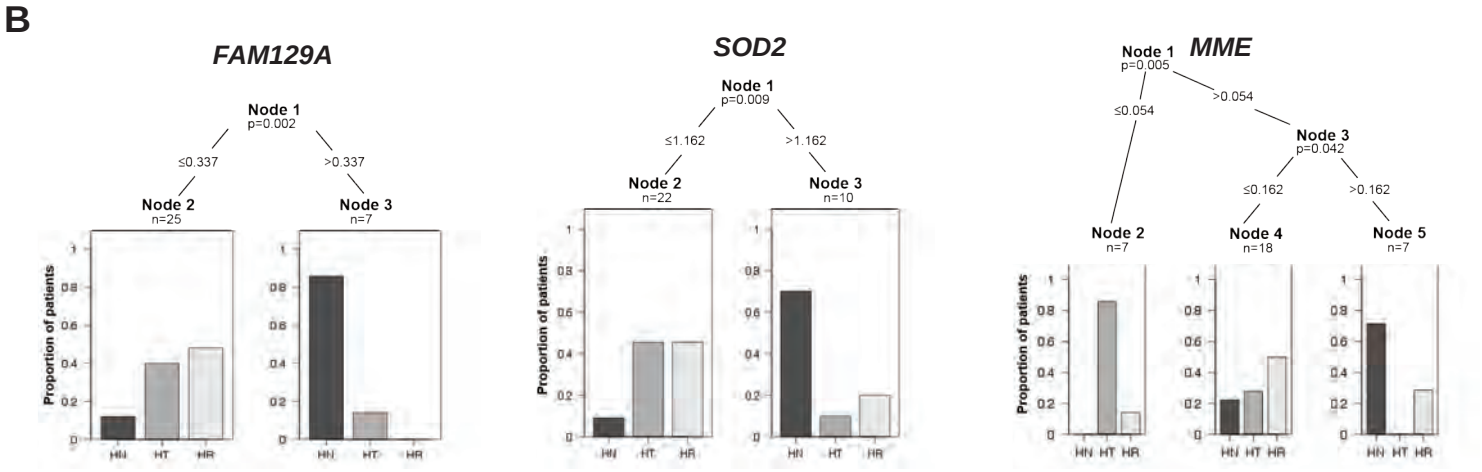

A
p<0.0001
FAM129A
p<0.0001
KRT7
p<0.0001
SOD2
p<0.0001
MME
log2 intensity
Control
Metastatic
Sample type

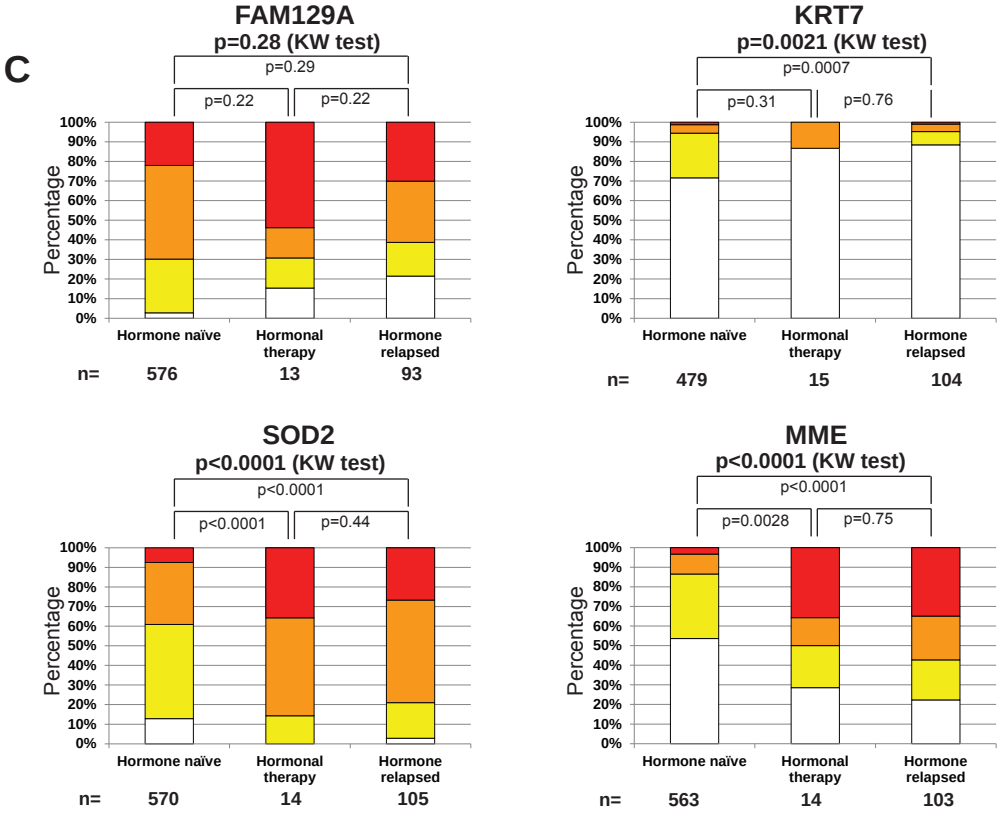

B
FAM129A
Benign
Tumour
p=0.17
Percentage
Benign
Metastatic
n= 12 10
KRT7
Benign
Tumour
p=0.0036
Percentage
Benign
Metastatic
n= 12 10
SOD2
Benign
Tumour
p=0.070
Percentage
Benign
Metastatic
n= 12 9
MME
Benign
Tumour
p<0.0001
Percentage
Benign
Metastatic
n= 11 9


C
MME (ng/ml)
p=0.72
Benign
Metastatic

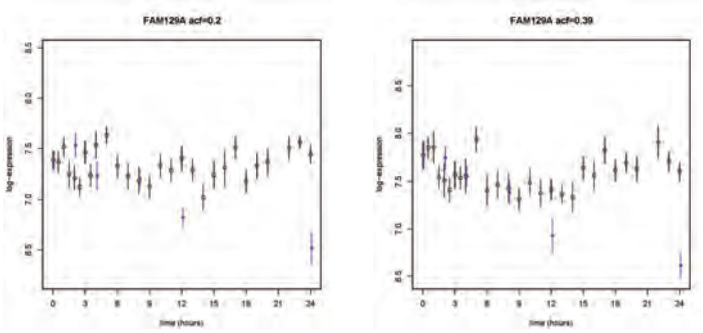

A
FAM129A
p=0.19
p=0.21
p=0.95
KRT7
p=0.15
p=0.0061
p=0.0019
SOD2
p=0.52
p=0.98
p=0.60
MME
p=0.32
p=0.16
p=0.011
Relative Expression
Low
Intermediate
High
B
KRT7
Node 1
p=0.027
≤0.00042
>0.00042
Node 2 (n=29)
Node 3 (n=18)
PSA
Node 1
p=0.031
≤11.3
>11.3
Node 2 (n=36)
Node 3 (n=13)
Low
Inter
High
Risk group
C
p=0.68
FAM129A
p<0.001
KRT7
p=0.018
SOD2
p=0.013
MME
log2 intensity
Control
Localised
Sample type
D
FAM129A
KRT7
SOD2
MME
Percentage
Benign
G3
G4
G5
n=
348
388
163
25
512
387
166
26
344
383
162
25
492
370
167
26
p<0.0001
Sensitivity - 69.5%
Specificity - 72.6%
p<0.0001
Sensitivity - 95.7%
Specificity - 45.3%
p<0.0001
Sensitivity - 61.0%
Specificity - 54.6%
p<0.0001
Sensitivity - 86.4%
Specificity - 75.1%


A
FAM129A
p=0.0056
p=0.0057
p=0.82
KRT7
p=0.014
p=0.17
p=0.17
SOD2
p=0.058
p=0.0057
p=0.16
MME
p=0.058
p=0.0031
p=0.090
Relative Expression
Hormone naive
Hormonal therapy
Hormone relapsed

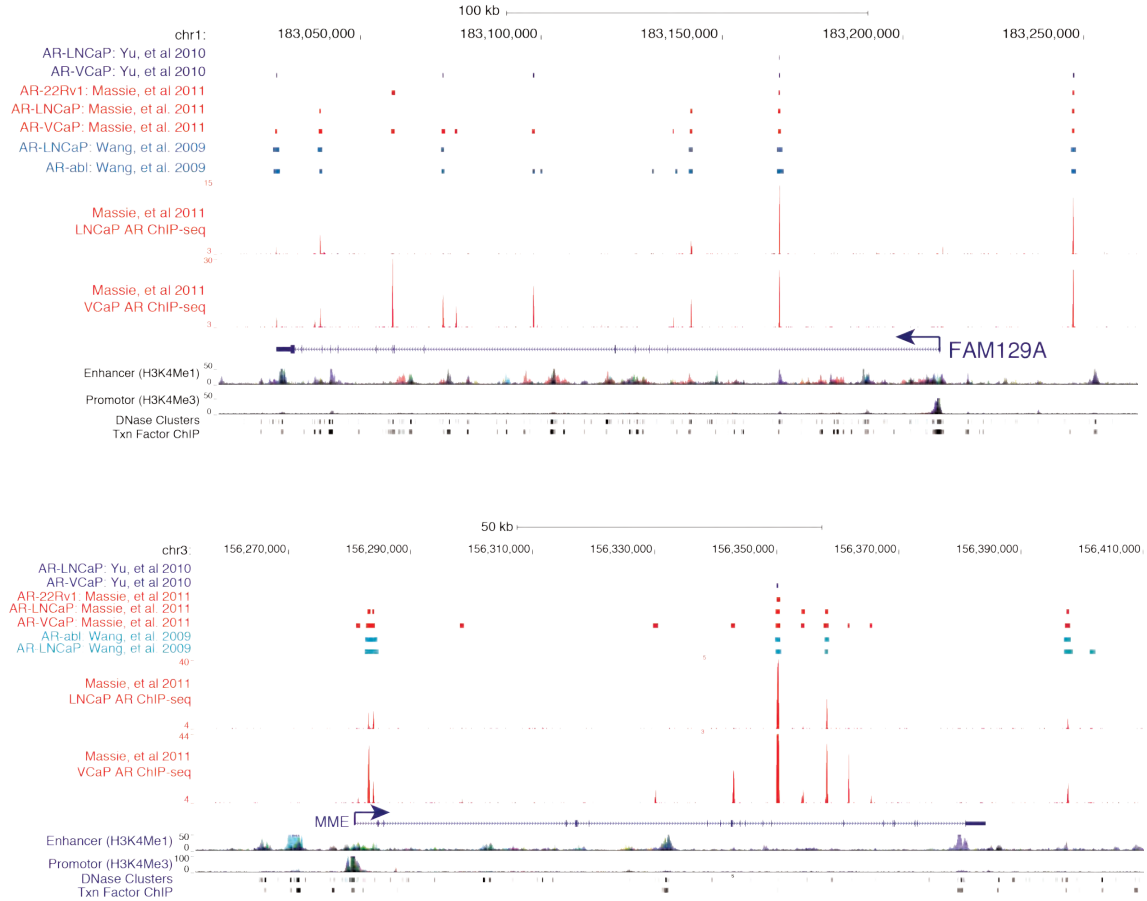

B
FAM129A
Node 1
p=0.002
≤0.337
>0.337
Node 2
n=25
Node 3
n=7
SOD2
Node 1
p=0.009
≤1.162
>1.162
Node 2
n=22
Node 3
n=10
MME
Node 1
p=0.005
≤0.054
>0.054
Node 3
p=0.042
≤0.162
>0.162
Node 2
n=7
Node 4
n=18
Node 5
n=7
Proportion of patients
HN
HT
HR


C
FAM129A
p=0.28 (KW test)
p=0.29
p=0.22
p=0.22
KRT7
p=0.0021 (KW test)
p=0.0007
p=0.31
p=0.76
SOD2
p<0.0001 (KW test)
p<0.0001
p<0.0001
p=0.44
MME
p<0.0001 (KW test)
p<0.0001
p=0.0028
p=0.75
Percentage
Hormone naïve
Hormonal therapy
Hormone relapsed
n= 576 13 93
n= 479 15 104
n= 570 14 105
n= 563 14 103

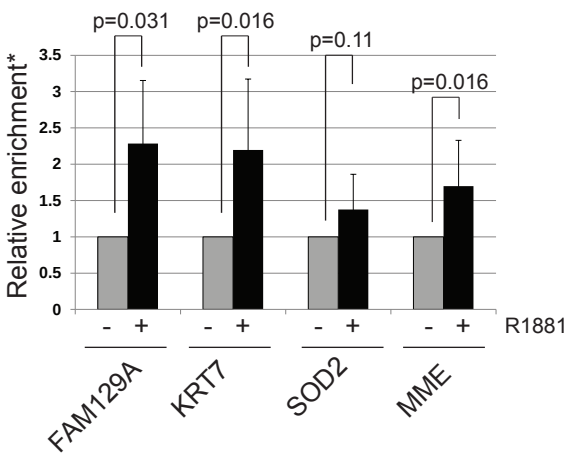
A

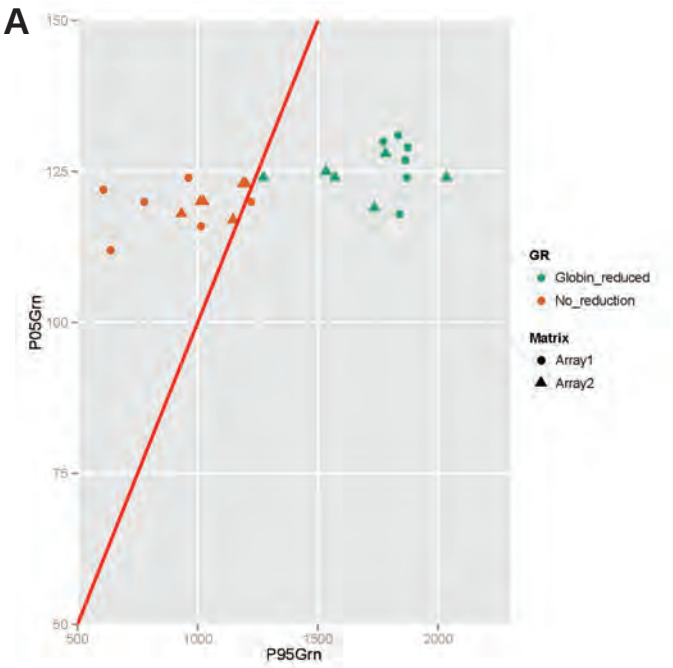

FAM129A acf=0.2
FAM129A acf=0.39

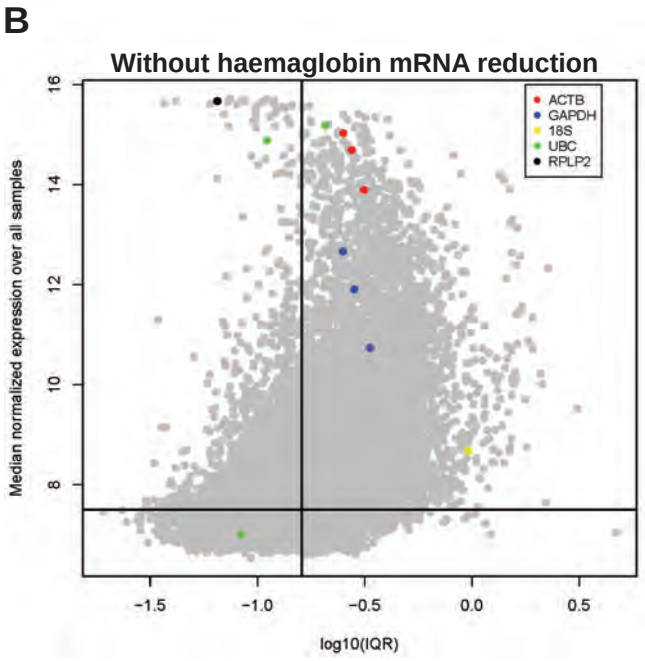
B

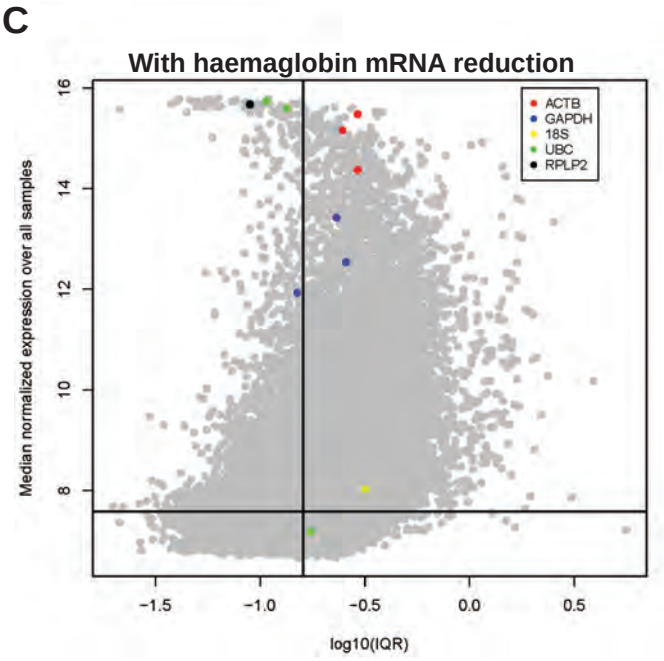

100 kb
chr1:
183,050,000
183,100,000
183,150,000
183,200,000
183,250,000
AR-LNCaP: Yu, et al 2010
AR-VCaP: Yu, et al 2010
AR-22Rv1: Massie, et al 2011
AR-LNCaP: Massie, et al. 2011
AR-VCaP: Massie, et al. 2011
AR-LNCaP: Wang, et al. 2009
AR-abl: Wang, et al. 2009
Massie, et al 2011
LNCaP AR ChIP-seq
Massie, et al 2011
VCaP AR ChIP-seq
FAM129A
Enhancer (H3K4Me1)
Promotor (H3K4Me3)
DNase Clusters
Txn Factor ChIP
50 kb
chr3:
156,270,000
156,290,000
156,310,000
156,330,000
156,350,000
156,370,000
156,390,000
156,410,000
AR-LNCaP: Yu, et al 2010
AR-VCaP: Yu, et al 2010
AR-22Rv1: Massie, et al 2011
AR-LNCaP: Massie, et al. 2011
AR-VCaP: Massie, et al. 2011
AR-abl: Wang, et al. 2009
AR-LNCaP: Wang, et al. 2009
Massie, et al 2011
LNCaP AR ChIP-seq
Massie, et al 2011
VCaP AR ChIP-seq
MME
Enhancer (H3K4Me1)
Promotor (H3K4Me3)
DNase Clusters
Txn Factor ChIP

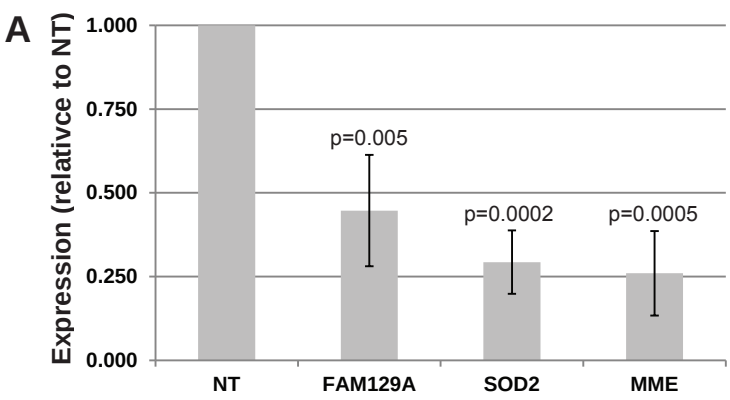
C

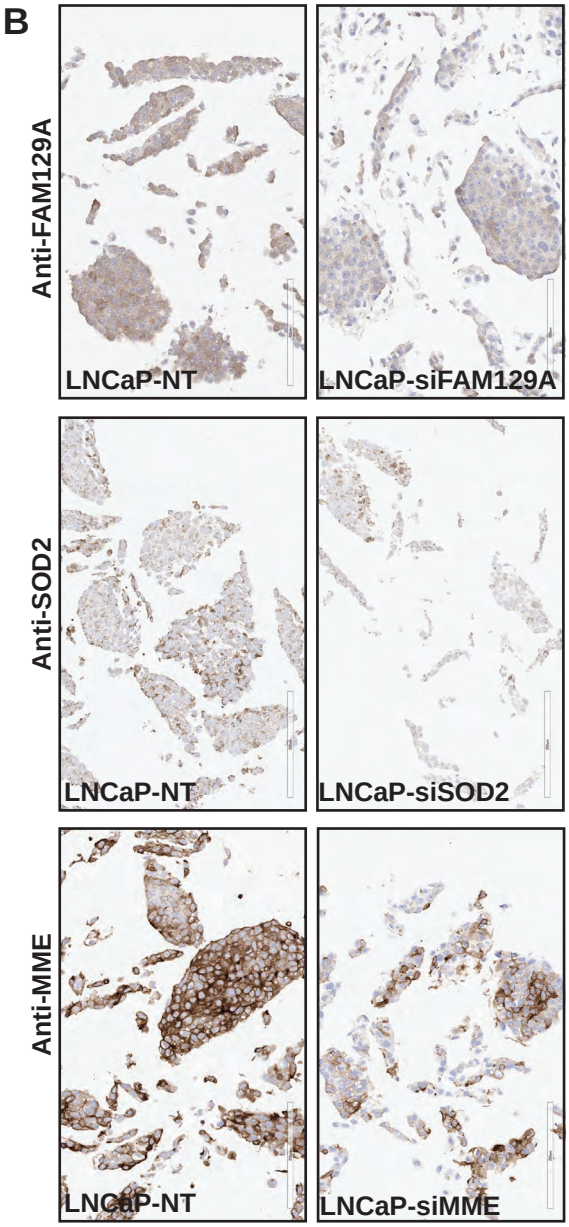

p=0.031
p=0.016
p=0.11
p=0.016
Relative enrichment*
3.5
3
2.5
2
1.5
1
0.5
0
- + - + - + - +
R1881
FAM129A
KRT7
SOD2
MME

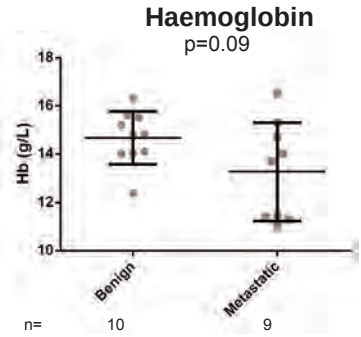

A
150
125
100
75
50
500
1000
1500
2000
P95Grn
P05Grn
GR
Globin_reduced
No_reduction
Matrix
Array1
Array2

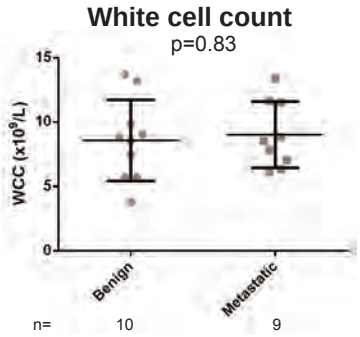

B
Without haemaglobin mRNA reduction
Median normalized expression over all samples
log10(IQR)
ACTB
GAPDH
18S
UBC
RPLP2

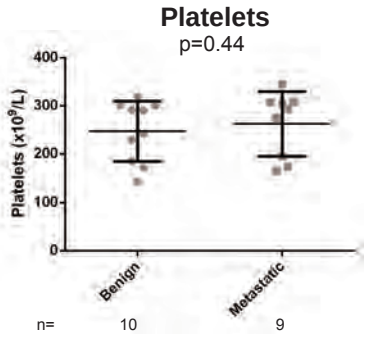

C
With haemaglobin mRNA reduction
Median normalized expression over all samples
log10(IQR)
ACTB
GAPDH
18S
UBC
RPLP2

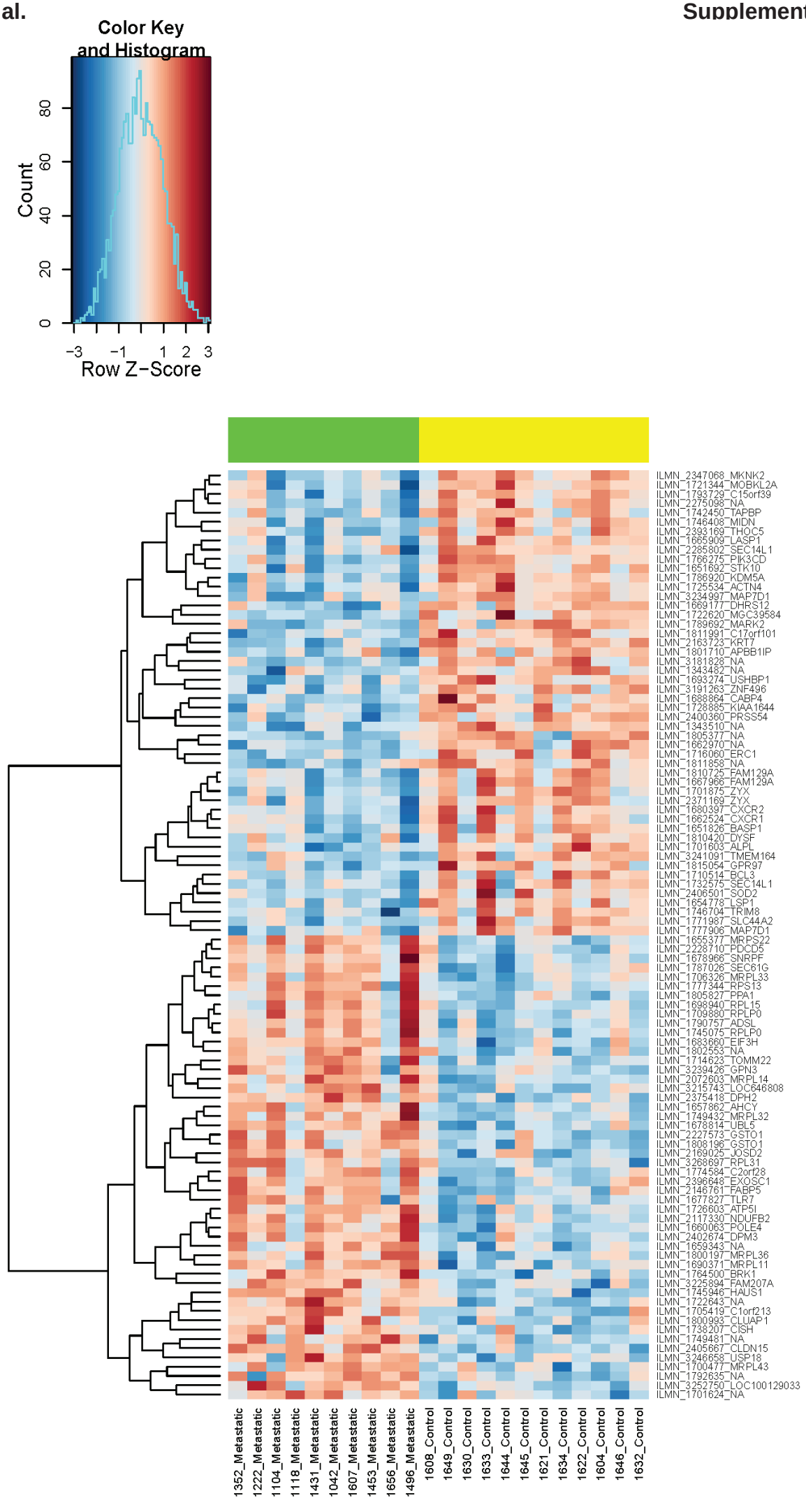

A
Expression (relativce to NT)
1.000
0.750
0.500
0.250
0.000
p=0.005
p=0.0002
p=0.0005
NT
FAM129A
SOD2
MME

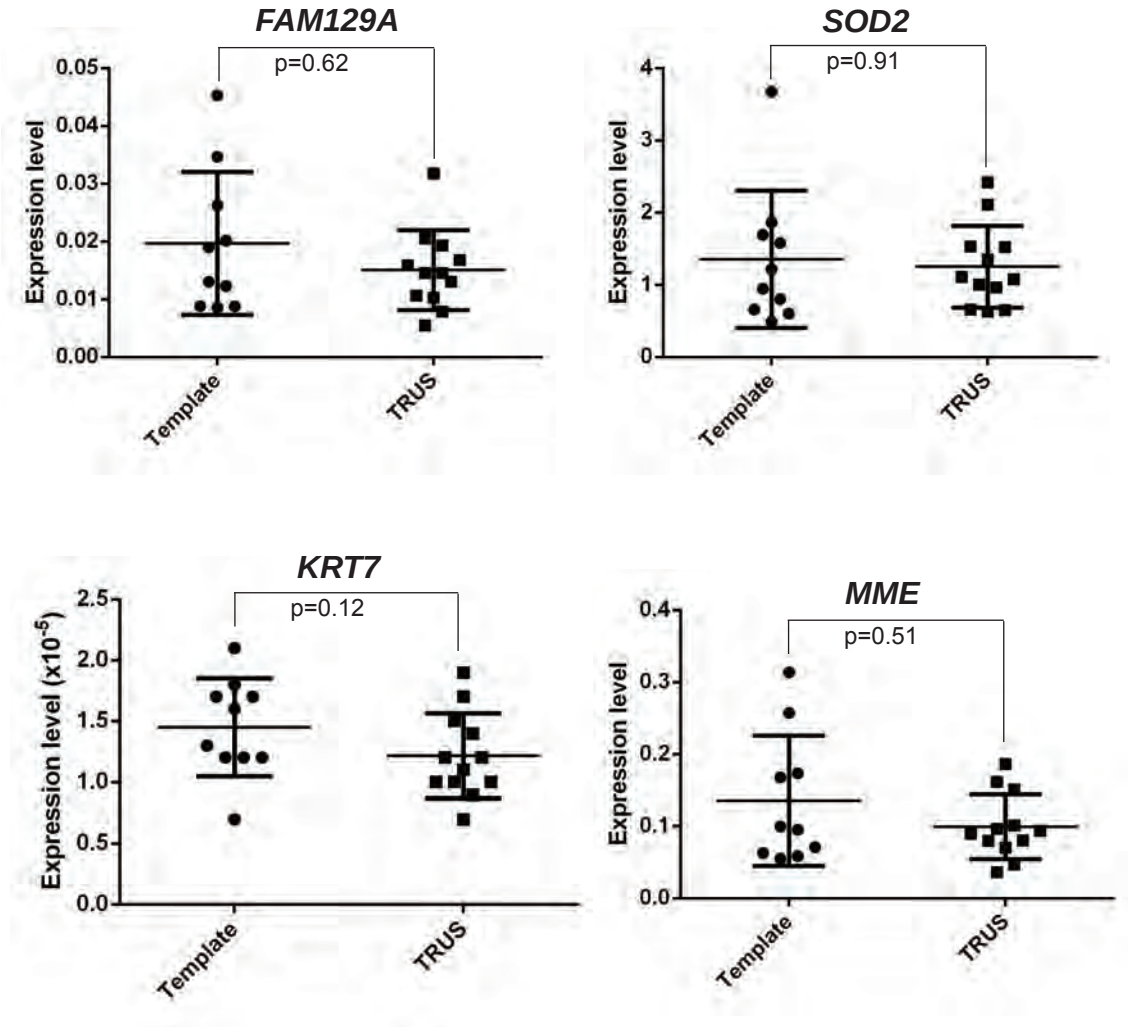

B
Anti-FAM129A
LNCaP-NT
LNCaP-siFAM129A
Anti-SOD2
LNCaP-NT
LNCaP-siSOD2
Anti-MME
LNCaP-NT
LNCaP-siMME


Haemoglobin
p=0.09
Hb (g/L)
18
16
14
12
10
Benign
Metastatic
n=
10
9

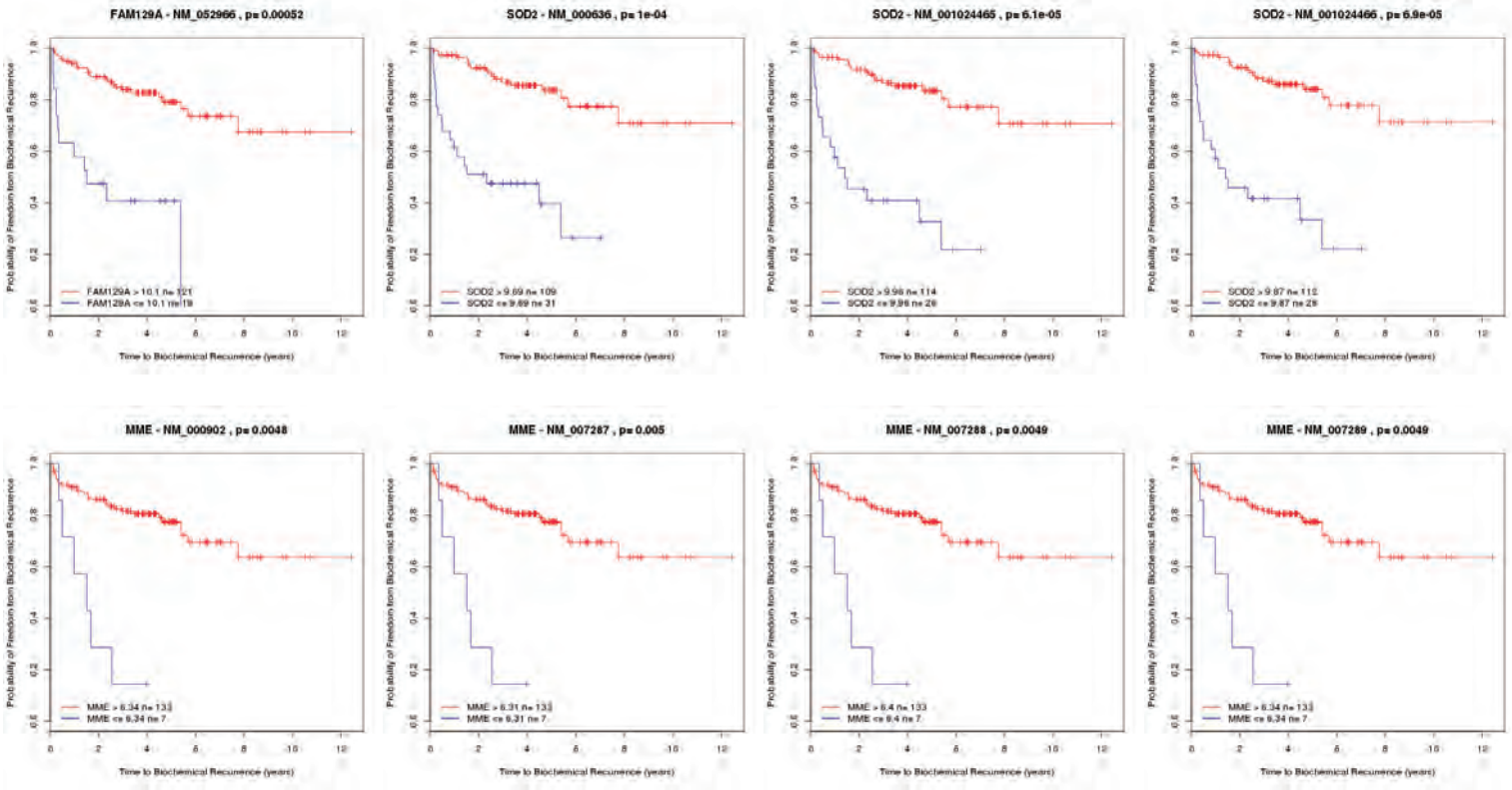

White cell count
p=0.83
WCC (x10⁹/L)
15
10
5
0
Benign
Metastatic
n=
10
9


Platelets
p=0.44
Platelets (x10⁹/L)
400
300
200
100
0
Benign
Metastatic
n=
10
9

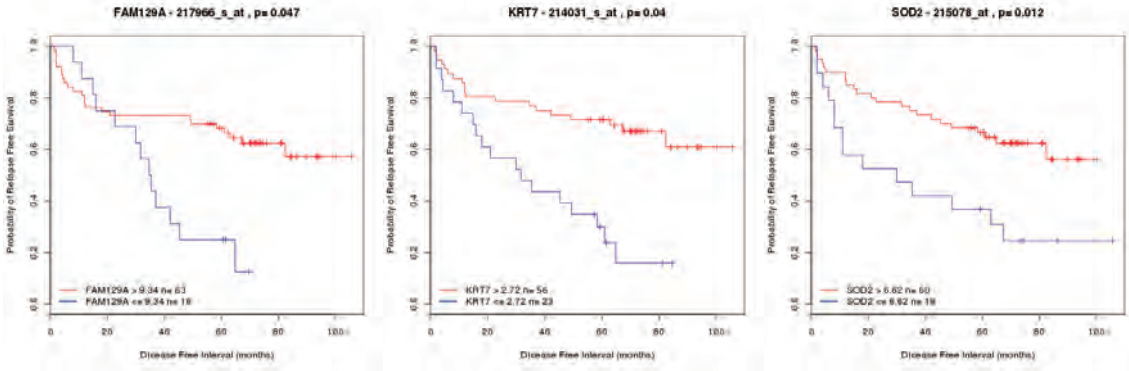

Color Key
and Histogram
Count
0
20
40
60
80
−3
−1
1
2
3
Row Z−Score
ILMN_2347068_MKNK2
ILMN_1721344_MOBKL2A
ILMN_1793729_C15orf39
ILMN_2275098_NA
ILMN_1742450_TAPBP
ILMN_1746408_MIDN
ILMN_2393169_THOC5
ILMN_1665909_LASP1
ILMN_2285802_SEC14L1
ILMN_1766275_PIK3CD
ILMN_1651692_STK10
ILMN_1786920_KDM5A
ILMN_1725534_ACTN4
ILMN_3234997_MAP7D1
ILMN_1669177_DHRS12
ILMN_1722620_MGC39584
ILMN_1789692_MARK2
ILMN_1811991_C17orf101
ILMN_2163723_KRT7
ILMN_1801710_APBB1IP
ILMN_3181828_NA
ILMN_1343482_NA
ILMN_1693274_USHBP1
ILMN_3191263_ZNF496
ILMN_1688864_CABP4
ILMN_1728885_KIAA1644
ILMN_2400380_PRSS54
ILMN_1343510_NA
ILMN_1805377_NA
ILMN_1662970_NA
ILMN_1716060_ERC1
ILMN_1811858_NA
ILMN_1810725_FAM129A
ILMN_1667966_FAM129A
ILMN_1701875_ZYX
ILMN_2371169_ZYX
ILMN_1680397_CXCR2
ILMN_1662524_CXCR1
ILMN_1651826_BASP1
ILMN_1810420_DYSF
ILMN_1701603_ALPL
ILMN_3241091_TMEM164
ILMN_1815054_GPR97
ILMN_1710514_BCL3
ILMN_1732575_SEC14L1
ILMN_2406501_SOD2
ILMN_1654778_LSP1
ILMN_1746704_TRIM8
ILMN_1771987_SLC44A2
ILMN_1777906_MAP7D1
ILMN_1655377_MRPS22
ILMN_2228710_PDCD5
ILMN_1678966_SNRPF
ILMN_1787026_SEC61G
ILMN_1706326_MRPL33
ILMN_1777344_RPS13
ILMN_1805827_PPA1
ILMN_1698940_RPL15
ILMN_1709880_RPLP0
ILMN_1790757_ADSL
ILMN_1745075_RPLP0
ILMN_1683660_EIF3H
ILMN_1802553_NA
ILMN_1714623_TOMM22
ILMN_3239426_GPN3
ILMN_2072603_MRPL14
ILMN_3215743_LOC646808
ILMN_2375418_DPH2
ILMN_1657862_AHCY
ILMN_1749432_MRPL32
ILMN_1678814_UBL5
ILMN_2227573_GSTO1
ILMN_1808196_GSTO1
ILMN_2169025_JOSD2
ILMN_3268697_RPL31
ILMN_1774584_C2orf28
ILMN_2396648_EXOSC1
ILMN_2146761_FABP5
ILMN_1677827_TLR7
ILMN_1726603_ATP5I
ILMN_2117330_NDUFB2
ILMN_1660063_POLE4
ILMN_2402674_DPM3
ILMN_1659343_NA
ILMN_1800197_MRPL36
ILMN_1690371_MRPL11
ILMN_1764500_BRK1
ILMN_3225894_FAM207A
ILMN_1745946_HAUS1
ILMN_1722643_NA
ILMN_1705419_C1orf213
ILMN_1800993_CLUAP1
ILMN_1738207_CISH
ILMN_1749481_NA
ILMN_2405667_CLDN15
ILMN_3246658_USP18
ILMN_1700477_MRPL43
ILMN_1792635_NA
ILMN_3252750_LOC100129033
ILMN_1701624_NA
1352_Metastatic
1222_Metastatic
1104_Metastatic
1118_Metastatic
1431_Metastatic
1042_Metastatic
1607_Metastatic
1453_Metastatic
1656_Metastatic
1496_Metastatic
1608_Control
1649_Control
1630_Control
1633_Control
1644_Control
1645_Control
1621_Control
1634_Control
1622_Control
1604_Control
1646_Control
1632_Control


FAM129A
p=0.62
Expression level
0.05
0.04
0.03
0.02
0.01
0.00
Template
TRUS
SOD2
p=0.91
Expression level
4
3
2
1
0
Template
TRUS
KRT7
p=0.12
Expression level (x10-5)
2.5
2.0
1.5
1.0
0.5
0.0
Template
TRUS
MME
p=0.51
Expression level
0.4
0.3
0.2
0.1
0.0
Template
TRUS

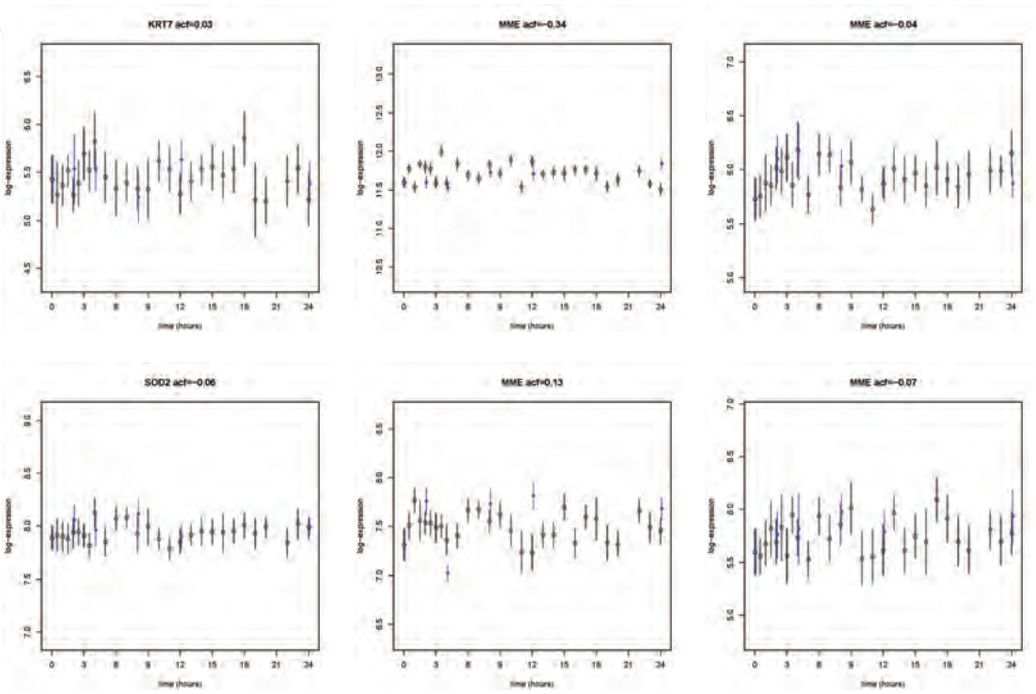
A


FAM129A - NM_052966 , p= 0.00052
SOD2 - NM_000636 , p= 1e-04
SOD2 - NM_001024465 , p= 6.1e-05
SOD2 - NM_001024466 , p= 6.9e-05
MME - NM_000902 , p= 0.0048
MME - NM_007287 , p= 0.005
MME - NM_007288 , p= 0.0049
MME - NM_007289 , p= 0.0049
Probability of Freedom from Biochemical Recurrence
Time to Biochemical Recurrence (years)

B


FAM129A - 217966_s_at , p= 0.047
KRT7 - 214031_s_at , p= 0.04
SOD2 - 215078_at , p= 0.012
Probability of Relapse Free Survival
Disease Free Interval (months)

A


KRT7 acf=0.03
MME acf=−0.34
MME acf=−0.04
SOD2 acf=−0.06
MME acf=0.13
MME acf=−0.07
log−expression
time (hours)

B

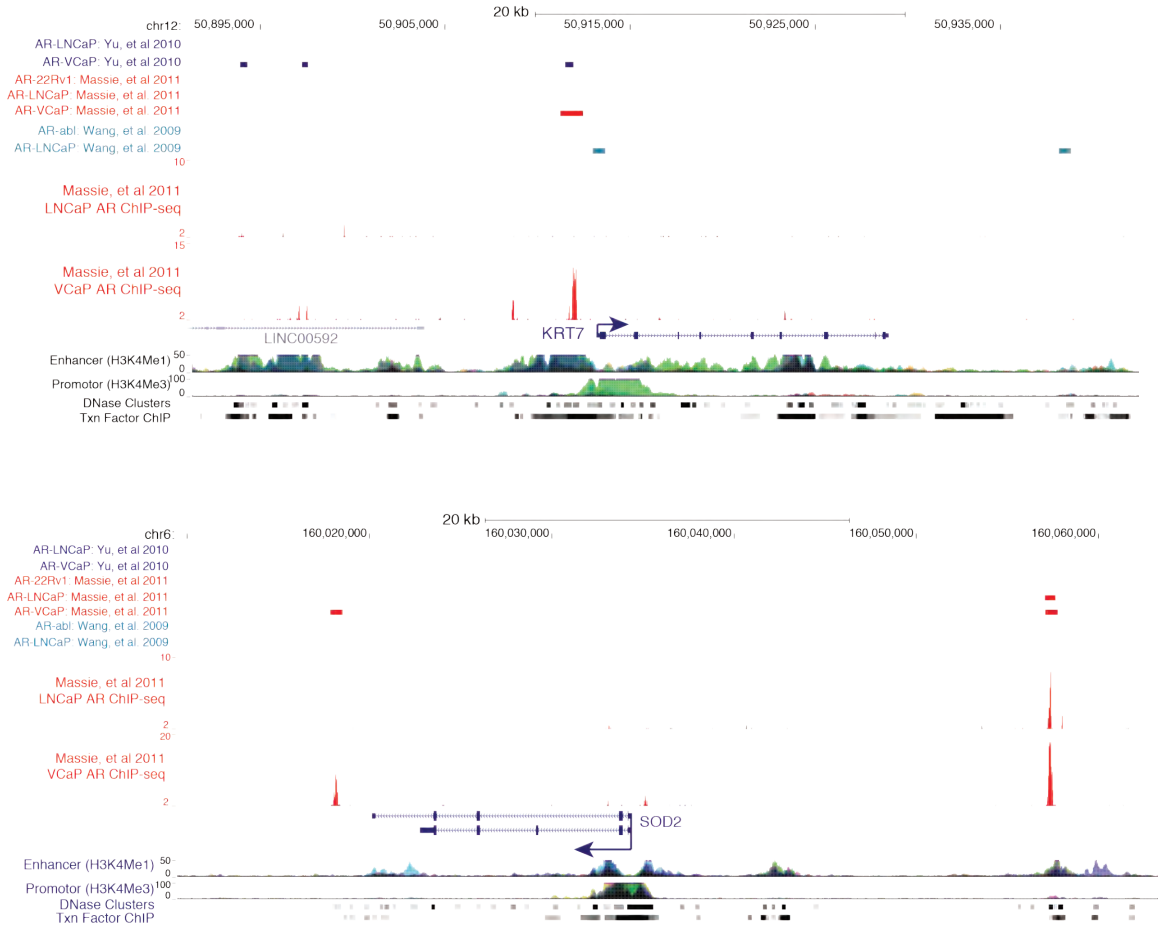

20 kb
chr12:
50,895,000
50,905,000
50,915,000
50,925,000
50,935,000
AR-LNCaP: Yu, et al 2010
AR-VCaP: Yu, et al 2010
AR-22Rv1: Massie, et al 2011
AR-LNCaP: Massie, et al. 2011
AR-VCaP: Massie, et al. 2011
AR-abl: Wang, et al. 2009
AR-LNCaP: Wang, et al. 2009
Massie, et al 2011
LNCaP AR ChIP-seq
Massie, et al 2011
VCaP AR ChIP-seq
LINC00592
KRT7
Enhancer (H3K4Me1)
Promotor (H3K4Me3)
DNase Clusters
Txn Factor ChIP
20 kb
chr6:
160,020,000
160,030,000
160,040,000
160,050,000
160,060,000
AR-LNCaP: Yu, et al 2010
AR-VCaP: Yu, et al 2010
AR-22Rv1: Massie, et al 2011
AR-LNCaP: Massie, et al. 2011
AR-VCaP: Massie, et al. 2011
AR-abl: Wang, et al. 2009
AR-LNCaP: Wang, et al. 2009
Massie, et al 2011
LNCaP AR ChIP-seq
Massie, et al 2011
VCaP AR ChIP-seq
SOD2
Enhancer (H3K4Me1)
Promotor (H3K4Me3)
DNase Clusters
Txn Factor ChIP

Table 1: Discovery cohort

Criteria	**Metastatic** (n=10)	**Benign TURP** (n=12)	**Benign template** (n=10)
Median age in years* (range)	65.0 (54.0-83.0)	66.5 (52.0-80.0)	61.0 (50.0-67.0)
Median PSA* (range)	43.0 (3.9-609.0)	7.7 (4.2-18.6)	10.8 (3.8-15.5)
Gleason sum score			
6 (%)	0/10 (0%)		
7 (%)	1/10 (20%)	N/A	N/A
≥8 (%)	9/10 (90%)	N/A	N/A
Clinical stage			
T3	7/10 (70%)	N/A	N/A
T4	3/10 (30%)	N/A	N/A
Hormone status			
Hormone naïve	6/10 (60%)	N/A	N/A
On hormone therapy	4/10 (40%)	N/A	N/A
Hormone refractory	0/10 (0%)	N/A	N/A

Table 2: Risk stratified validation cohort

Criteria	**Low risk** (n=12)	**Moderate risk** (n=21)	**High risk** (n=14)
Median age in years* (range)	61.0 (51.0-71.0)	66.0 (53.0-87.0)	69.0 (55.0-79.0)
Median PSA* (range)	5.7 (2.7-10.0)	9.0 (2.1-18.5)	14.0 (4.0-149.0)
Gleason score			
6 (%)	12/12 (100%)	0/21 (0%)	1/14 (7%)
7 (%)	0/12 (0%)	21/21 (100%)	1/14 (7%)
≥8 (%)	0/12 (0%)	0/21 (0%)	12/14 (86%)
Clinical stage			
T2	12/12 (100%)	19/21 (91%)	7/14 (50%)
T3	0/12 (0%)	2/21 (9%)	7/14 (50%)
T4	0/12 (0%)	0/21 (0%)	0/14 (0%)
+ve biopsy cores (%)	21% (6%-67%)	31% (8%-73%)	78% (17%-100%)

Table 3: Defined hormone status validation cohort

Criteria	**Hormone naïve** (n=9)	**Stable disease** (n=12)	**Hormone refractory** (n=12)
Median age in years* (range)	75.0 (64.0-93.0)	66.5 (61.0-88.0)	71.0 (51.0-82.0)
Median PSA* (range)	123.0 (12.5-511.6)	85.0 (7.4-6470.0)	41.1 (2.0-405.0)
Gleason sum score			
No biopsy taken, diagnosis based on clinical criteria	3/9 (33%)	6/12 (50%)	4/12 (34%)
6	0/10 (0%)	1/12 (8%)	1/12 (8%)
7	1/10 (11%)	3/12 (25%)	1/12 (8%)
≥8	5/10 (56%)	2/12 (17%)	6/12 (50%)
Clinical stage			
T2	0/9 (0%)	1/12 (8%)	1/12 (8%)
T3	7/9 (77%)	9/12 (75%)	8/12 (67%)
T4	3/10 (33%)	2/12 (17%)	3/12 (25%)

Table 4: Hormone relapse TMA

Treatment	Number of patients (%)
LHRH, no progression (hormonal therapy cohort)	9 (15.2%)
Relapse on LHRH	19 (32.2%)
Relapse on complete androgen blockade	18 (30.5%)
Relapse following external beam radiation	9 (15.2%)
Salvage TURP post relapse	4 (6.8%)

Table 5:

	Discovery Cohort (Benign vs metastatic)			**Defined Hormone Status Cohort (HN vs HR)**		**Risk Stratified Localised Cohort**		
	PAXgene (Expression/qPCR)	*Tissue mRNA*	*Tissue protein (biopsy)*	*PAXgene*	*Tissue protein*	*PAXgene (inter vs high)*	*Tissue mRNA (benign vs tumour)*	*Tissue protein (benign vs tumour)*
FAM129A	↓/↓	↓	↔	↓	↔	↔	↔	↑
KRT7	↓/↔	↓	↓	↓	↓	↓	↓	↓
SOD2	↓/↓	↓	↔	↔	↑	↔	↑	↓
MME	↓/↓	↓	↓	↔	↑	↓	↓	↓

Gene name	Aliases	Function	Role in any cancer	Known role in prostate cancer	References
FAM129A	Niban, C1orf24	Stress induced. Regulates eIF2alpha and S6K1-BP1. Regulates p53 via nucleophosmin and MDM2.	Overexpressed in cancers of head and neck, thyroid, renal.	None	(Adachi, et al. 2004; Cerutti, et al. 2006; Ito, et al. 2010; Ji, et al. 2012; Matsumoto, et al. 2006; Patel, et al. 2011; Sun, et al. 2007)
PPA1	PP1	Catalyzes the hydrolysis of pyrophosphate to inorganic phosphate. Regulates neurite growth.	Regulates cell migration and invasion. Overexpressed in hepatocellular, gastric and ovarian cancers.	None	(Jeong, et al. 2012; Megger, et al. 2013; Wang, et al. 2012)
DPM3	Prostin-1	Acts as a stabilizer subunit of the dolichyl-phosphate mannosyltransferase complex.	None	Negatively associated with prostate tumour invasion.	(Ashida, et al. 2006; Manos, et al. 2001)
SOD2	Manganese Containing Superoxide Dismutase	Binds to the superoxide by products of oxidative phosphorylation and converts them to hydrogen peroxide and oxygen.	Promotes metastases in multiple tumour types.	Down regulated in prostate cancer. Raised in circulating prostate tumour cells. SOD2 mimetics act as AR inhibitors in HRPC cancer.	(Bostwick, et al. 2000; Giesing, et al. 2012; Hempel, et al. 2011; Thomas and Sharifi 2012)
KRT7	Cytokeratin -7	Expressed during differentiation of simple and stratified epithelial tissues	Used in many cancer types as a basal cell marker or to distinguish metastatic from primary disease	Used as a basal cell marker for diagnosis.	(Goldstein 2002; Tot 2002; Yaziji, et al. 2001)
MME	CD10	A neutral endopeptidase that cleaves peptides and inactivates several peptide hormones	Overexpressed in many tumour types including renal, ovarian and pancreatic cancers.	Expression is linked to aggressive disease and poor outcome.	(Fleischmann, et al. 2008; Ho, et al. 2013; Maguer-Satta, et al. 2011; Voutsadakis, et al. 2012)

Supplementary Table 1:

IlluminaID	SYMBOL	logFC	t	P.Value	adj.P.Val	B	Taylor/Sawyers significant	Glinsky significant	Ab quality	Ab source/	Comments
ILMN_1667966	FAM129A	7.05E-01	5.42E+00	9.15E-06	0.2297948	2.56199472	5.20E-04	0.047	****	HPA	Strong staining
ILMN_2405667	CLDN15	-2.58E-01	-5.41E+00	9.55E-06	0.2297948	2.53127346	1.50E-02	0.0057	***	HPA	Moderate staining, no discrimination betw een tumour/normal
ILMN_2163723	KRT7	2.43E-01	5.18E+00	1.78E-05	0.284781	2.08562967	1.50E-01	0.04	*****	HPA/ Novocastra	Strong staining
ILMN_1777344	RPS13	-4.53E-01	-4.89E+00	3.85E-05	0.3512191	1.52292897	8.50E-01	NA	*	HPA	No expression
ILMN_1660063	POLE4	-4.21E-01	-4.74E+00	5.82E-05	0.3512191	1.22062321	4.70E-01	NA	N/A	HPA	None available
ILMN_2402674	DPM3	-3.90E-01	-4.74E+00	5.88E-05	0.3512191	1.21376562	2.30E-02	0.62	****	HPA	Strong staining
ILMN_2117330	NDUFB2	-4.11E-01	-4.69E+00	6.62E-05	0.3512191	1.12636348	7.50E-01	0.46	*	HPA	Very w eak
ILMN_1657862	AHCY	-3.29E-01	-4.62E+00	8.11E-05	0.3512191	0.97703084	6.80E-02	0.32	*	HPA	Weak and non-specific staining
ILMN_1745946	HAUS1	-2.25E-01	-4.60E+00	8.55E-05	0.3512191	0.93735357	1.20E-01	NA	**	HPA	Moderate staining, lots of non-specific staining in muscle
ILMN_1701603	ALPL	7.61E-01	4.50E+00	1.10E-04	0.3512191	0.74858287	1.20E-01	0.16	**	HPA	Weak/moderate staining
ILMN_2347068	MKNK2	4.20E-01	4.50E+00	1.11E-04	0.3512191	0.74698355	2.60E-01	0.97	**	HPA	Weak/moderate staining
ILMN_1654778	LSP1	3.73E-01	4.47E+00	1.21E-04	0.3512191	0.67959336	1.00E-01	0.059	*	HPA	No epithelial expression
ILMN_1726603	ATP5I	-3.33E-01	-4.43E+00	1.35E-04	0.3512191	0.60067172	5.40E-01	0.21	*****	HPA	Strong staining
ILMN_1805827	PPA1	-5.49E-01	-4.37E+00	1.60E-04	0.3609522	0.47498633	8.90E-01	0.022	****	HPA	Moderate staining, excellent discrimination benign/cancer
ILMN_3241091	TMEM164	6.66E-01	4.36E+00	1.62E-04	0.3609522	0.46532921	3.80E-01	0.13	**	HPA	Weak, lots of non-specific
ILMN_1678814	UBL5	-2.93E-01	-4.33E+00	1.75E-04	0.3609522	0.40800261	9.20E-01	0.00005	N/A	HPA	None available
ILMN_1749432	MRPL32	-2.86E-01	-4.30E+00	1.94E-04	0.3609522	0.33075996	1.10E-01	NA	*	HPA	No/w eak staining
ILMN_1709880	RPLP0	-5.10E-01	-4.25E+00	2.19E-04	0.3609522	0.2412398	1.50E-01	0.29	**	HPA	Weak/moderate staining
ILMN_1786920	KDM5A	3.43E-01	4.22E+00	2.36E-04	0.3609522	0.18506283	1.70E-02	0.11	***	HPA	Moderate staining, no discrimination betw een tumour/normal
ILMN_1701875	ZYX	4.21E-01	4.21E+00	2.46E-04	0.3609522	0.15170055	3.20E-01	0.0078	*	HPA	Very w eak
ILMN_1669177	DHRS12	2.29E-01	4.20E+00	2.50E-04	0.3609522	0.13987586	3.50E-01	0.053	**	HPA	Weak/moderate staining
ILMN_2227573	GSTO1	-3.14E-01	-4.20E+00	2.52E-04	0.3609522	0.13442907	9.10E-01	0.38	*	HPA	No/w eak staining
ILMN_1693274	USHBP1	2.08E-01	4.19E+00	2.57E-04	0.3609522	0.12001158	1.50E-01	NA	**	HPA	Weak/moderate staining
ILMN_2406501	SOD2	6.79E-01	4.17E+00	2.74E-04	0.3609522	0.07258804	6.10E-05	NA	*****	HPA	Strong staining, some discrimination
ILMN_2400360	PRSS54	2.55E-01	4.16E+00	2.81E-04	0.3609522	0.05338254	5.70E-02	NA	***	HPA	Moderate staining, possible discrimination betw een tumour/normal
ILMN_1793729	C15orf39	4.30E-01	4.14E+00	2.92E-04	0.3609522	0.02520559	NA	NA	*	HPA	No/w eak staining
ILMN_3239426	GPN3	-2.94E-01	-4.14E+00	2.92E-04	0.3609522	0.02476414	9.20E-01	0.81	*	HPA	No/w eak staining
ILMN_1714623	TOMM22	-2.88E-01	-4.12E+00	3.07E-04	0.3609522	-0.01373511	2.70E-01	0.11	*****	HPA	Strong staining
ILMN_1766275	PIK3CD	2.52E-01	4.12E+00	3.15E-04	0.3609522	-0.03236825	4.60E-02	0.14	*	HPA	No/w eak staining
ILMN_1746408	MIDN	3.76E-01	4.10E+00	3.29E-04	0.3609522	-0.06343384	2.80E-01	NA	*	HPA	No/w eak staining
ILMN_1789692	MARK2	2.20E-01	4.09E+00	3.35E-04	0.3609522	-0.07715236	4.60E-01	0.039	***	HPA	Moderate staining
ILMN_1728885	KIAA1644	2.15E-01	4.09E+00	3.41E-04	0.3609522	-0.09024041	4.50E-01	0.058	***	HPA	Moderate staining
ILMN_1651826	BASP1	5.03E-01	4.07E+00	3.56E-04	0.3609522	-0.12301803	7.50E-01	0.053	*****	HPA	Strong staining, some discrimination
ILMN_2169025	JOSD2	-2.51E-01	-4.06E+00	3.64E-04	0.3609522	-0.14096234	1.20E-01	NA	N/A	HPA	None available
ILMN_2146761	FABP5	-2.83E-01	-4.03E+00	3.99E-04	0.3609522	-0.20925884	9.40E-01	0.82	****	HPA	Strong staining in a small proportion of cancers, others negative
ILMN_1725534	ACTN4	3.26E-01	4.03E+00	4.00E-04	0.3609522	-0.21099725	6.10E-02	0.19	*****	HPA	Strong staining
ILMN_1746704	TRIM8	2.73E-01	4.02E+00	4.01E-04	0.3609522	-0.21309143	9.90E-01	0.14	*	HPA	No/w eak staining
ILMN_3246658	USP18	-4.43E-01	-4.02E+00	4.02E-04	0.3609522	-0.2151573	9.00E-01	0.63	*	HPA	No/w eak staining, strong staining in muscle
ILMN_1680397	CXCR2	6.18E-01	4.01E+00	4.14E-04	0.3609522	-0.2364445	3.50E-02	0.66	*	HPA	No/w eak staining
ILMN_1738207	CISH	-2.62E-01	-4.01E+00	4.17E-04	0.3609522	-0.24219118	8.00E-01	0.019	***	HPA	Moderate staining
ILMN_1688864	CABP4	2.44E-01	4.01E+00	4.18E-04	0.3609522	-0.24362427	1.80E-01	NA	*	HPA	No/w eak staining
ILMN_2072603	MRPL14	-2.60E-01	-3.99E+00	4.43E-04	0.3609522	-0.28759452	5.60E-04	NA	***	HPA	Moderate staining, possible discrimination betw een tumour/normal
ILMN_1810420	DYSF	5.90E-01	3.98E+00	4.49E-04	0.3609522	-0.29798572	7.80E-02	0.75	*	HPA	No/w eak staining
ILMN_1698940	RPL15	-3.49E-01	-3.98E+00	4.52E-04	0.3609522	-0.30157664	1.70E-01	0.0058	*	HPA	No/w eak staining
ILMN_1716060	ERC1	2.41E-01	3.98E+00	4.56E-04	0.3609522	-0.30839306	7.50E-04	0.034	***	HPA	Moderate/strong staining, various localisations
ILMN_1764500	BRK1	-2.40E-01	-3.97E+00	4.60E-04	0.3609522	-0.31617082	1.00E-01	NA	N/A	HPA	None available
ILMN_1705419	C1orf213	-2.01E-01	-3.97E+00	4.65E-04	0.3609522	-0.32349844	NA	NA	*	HPA	No/w eak staining
ILMN_2375418	DPH2	-2.07E-01	-3.97E+00	4.69E-04	0.3609522	-0.32927813	2.20E-02	0.1	*	HPA	No/w eak staining

IlluminaID	SYMBOL	logFC	t	P.Value	adj.P.Val	B	Taylor/Sawyers significant	Glinsky significant	Ab quality	Ab source/	Comments
ILMN_1655377	MRPS22	-2.83E-01	-3.95E+00	4.90E-04	0.3609522	-0.36360326	2.90E-01	0.75	***	HPA	Moderate staining
ILMN_3215743	LOC646808	-2.11E-01	-3.95E+00	4.91E-04	0.3609522	-0.36512181	NA	NA	N/A	HPA	None available
ILMN_2228710	PDCD5	-3.36E-01	-3.95E+00	4.94E-04	0.3609522	-0.36874949	8.80E-02	0.54	*****	HPA	Strong staining, some discrimination
ILMN_3252750	LOC100129033	-1.88E-01	-3.95E+00	4.94E-04	0.3609522	-0.3692881	NA	NA	N/A	HPA	None available
ILMN_1706326	MRPL33	-3.38E-01	-3.95E+00	4.95E-04	0.3609522	-0.37087403	4.50E-01	0.99	N/A	HPA	None available
ILMN_3268697	RPL31	-4.00E-01	-3.93E+00	5.17E-04	0.3662037	-0.40350473	2.90E-01	0.025	N/A	HPA	None available
ILMN_3191263	ZNF496	2.44E-01	3.91E+00	5.44E-04	0.3665536	-0.44128924	1.10E-01	NA	*	HPA	No/w eak staining
ILMN_1801710	APBB1IP	3.93E-01	3.91E+00	5.49E-04	0.3665536	-0.44765023	5.90E-01	0.53	*	HPA	No/w eak staining
ILMN_1651692	STK10	2.72E-01	3.91E+00	5.50E-04	0.3665536	-0.44913407	1.90E-01	0.69	*	HPA	No/w eak staining
ILMN_1777906	MAP7D1	3.24E-01	3.90E+00	5.56E-04	0.3665536	-0.45815583	4.70E-01	0.0004	*	HPA	No/w eak staining
ILMN_1732575	SEC14L1	4.28E-01	3.89E+00	5.70E-04	0.3692906	-0.47668911	6.30E-01	0.057	*****	HPA	Strong staining
ILMN_1710514	BCL3	4.01E-01	3.89E+00	5.76E-04	0.3692906	-0.48405835	7.60E-02	0.033	**	HPA	Weak/moderate staining
ILMN_1721344	MOBKL2A	3.66E-01	3.87E+00	5.99E-04	0.3790569	-0.5136351	NA	NA	*	HPA	No/w eak staining
ILMN_1800993	CLUAP1	-2.22E-01	-3.86E+00	6.14E-04	0.3836048	-0.53242992	2.00E-01	0.09	***	HPA	Moderate/strong staining, no discrimination
ILMN_2396648	EXOSC1	-2.55E-01	-3.84E+00	6.52E-04	0.3892703	-0.57784993	2.90E-02	0.3	***	HPA	Moderate staining
ILMN_1722620	MGC39584	1.78E-01	3.84E+00	6.64E-04	0.3892703	-0.59082668	NA	NA	N/A	HPA	None available
ILMN_1683660	EIF3H	-2.64E-01	-3.84E+00	6.64E-04	0.3892703	-0.59108285	2.20E-01	0.33	***	HPA	Moderate staining, possible discrimination betw een tumour/normal
ILMN_3225894	FAM207A	-2.70E-01	-3.83E+00	6.74E-04	0.3892703	-0.60285284	8.90E-03	NA	*****	HPA	Strong staining, no discrimination
ILMN_1790757	ADSL	-2.41E-01	-3.82E+00	6.85E-04	0.3892703	-0.61459751	3.00E-01	0.043	***	HPA	Moderate/strong staining, no discrimination
ILMN_1678966	SNRPF	-3.33E-01	-3.82E+00	6.86E-04	0.3892703	-0.61578841	1.50E-01	0.068	N/A	HPA	None available
ILMN_1815054	GPR97	4.01E-01	3.82E+00	6.91E-04	0.3892703	-0.62119907	3.00E-02	0.58	***	HPA	Moderate staining
ILMN_1665909	LASP1	2.96E-01	3.80E+00	7.38E-04	0.3892703	-0.671272	8.60E-01	0.12	***	HPA	Moderate staining
ILMN_1700477	MRPL43	-2.21E-01	-3.79E+00	7.43E-04	0.3892703	-0.67569775	6.20E-01	NA	N/A	HPA	None available
ILMN_1811991	C17orf101	1.98E-01	3.79E+00	7.45E-04	0.3892703	-0.6774734	NA	NA	*	HPA	No/w eak staining
ILMN_1771987	SLC44A2	2.68E-01	3.79E+00	7.48E-04	0.3892703	-0.68142212	3.50E-01	NA	*	HPA	No/w eak staining
ILMN_1690371	MRPL11	-2.43E-01	-3.79E+00	7.51E-04	0.3892703	-0.68371705	1.70E-02	0.59	***	HPA	Weak/moderate staining, possible discrimination betw een tumour/normal
ILMN_1677827	TLR7	-3.38E-01	-3.79E+00	7.52E-04	0.3892703	-0.68544161	9.70E-03	0.48	*	HPA	No/w eak staining
ILMN_2393169	THOC5	2.71E-01	3.79E+00	7.54E-04	0.3892703	-0.68696011	4.10E-01	0.21	*	HPA	No/w eak staining
ILMN_1800197	MRPL36	-2.61E-01	-3.78E+00	7.62E-04	0.3892703	-0.69538763	1.70E-01	NA	***	HPA	Moderate staining, possible discrimination betw een tumour/normal
ILMN_1742450	TAPBP	3.54E-01	3.76E+00	8.02E-04	0.3892703	-0.73366022	1.50E-01	0.58	***	HPA	Moderate staining
ILMN_1662524	CXCR1	5.68E-01	3.76E+00	8.15E-04	0.3892703	-0.74520579	1.70E-02	0.43	*	HPA	No/w eak staining
ILMN_1774584	C2orf28	-2.59E-01	-3.75E+00	8.21E-04	0.3892703	-0.75152148	NA	NA	*	HPA	No/w eak staining
ILMN_1787026	SEC61G	-3.74E-01	-3.75E+00	8.40E-04	0.3892703	-0.76854983	3.80E-01	0.018	*	HPA	No/w eak staining
ILMN_2067370	SNRPF	-3.89E-01	-3.73E+00	8.69E-04	0.3892703	-0.7942236	1.50E-01	6.80E-02	N/A	HPA	None available
ILMN_1753912	ABHD11	1.98E-01	3.73E+00	8.73E-04	0.3892703	-0.79751531	8.10E-03	0.28	***	HPA	Moderate staining, possible discrimination betw een tumour/normal
ILMN_3307772	PACSIN2	3.37E-01	3.73E+00	8.86E-04	0.3892703	-0.80875555	8.00E-03	0.64	***	HPA	Moderate/strong staining
ILMN_1660847	PFKFB3	2.80E-01	3.72E+00	9.06E-04	0.3892703	-0.8254048	2.00E-02	0.53	***	HPA/SDIX	Moderate staining
ILMN_2262362	ABCG1	2.61E-01	3.72E+00	9.10E-04	0.3892703	-0.82819301	8.90E-01	1.60E-01	*	HPA	No/w eak staining
ILMN_1692707	PTRHD1	-2.17E-01	-3.71E+00	9.16E-04	0.3892703	-0.83370526	4.00E-01	NA	**	HPA	Weak/moderate staining
ILMN_2100815	TMEM9B	-2.20E-01	-3.71E+00	9.34E-04	0.3892703	-0.84824095	2.70E-01	4.40E-01	***	HPA	Moderate staining, lots of non-specific staining
ILMN_1696554	UBE2Q2P1	-1.81E-01	-3.71E+00	9.36E-04	0.3892703	-0.85007213	1.40E-01	NA	N/A	HPA	None available
ILMN_3290681	OGFOD1	2.18E-01	3.70E+00	9.38E-04	0.3892703	-0.85143553	1.00E-01	8.60E-01	**	HPA	Weak/moderate staining
ILMN_2391765	C6orf48	-5.03E-01	-3.70E+00	9.39E-04	0.3892703	-0.85243815	NA	NA	N/A	HPA	None available
ILMN_1669033	NCOA1	3.60E-01	3.70E+00	9.45E-04	0.3892703	-0.85717582	7.00E-01	1.60E-01	*****	HPA/	Moderate/strong staining, possible discrimination betw een tumour/normal
										Cell signalling Technology	
ILMN_1739497	GTF2H5	-2.03E-01	-3.70E+00	9.52E-04	0.3892703	-0.86239138	1.10E-01	3.00E-01	***	HPA/	Moderate staining
										Sigma Aldrich	
ILMN_2275396	ADORA1	-1.96E-01	-3.70E+00	9.54E-04	0.3892703	-0.86424815	2.40E-02	7.00E-01	N/A	HPA	None available
ILMN_2206716	JTB	-2.01E-01	-3.69E+00	9.83E-04	0.3892703	-0.88660582	1.20E-01	3.00E-01	**	HPA	Weak/moderate staining
ILMN_1789793	NUAK2	2.64E-01	3.68E+00	9.91E-04	0.3892703	-0.89283829	6.40E-01	NA	***	HPA	Moderate staining
ILMN_2400183	DPH5	-2.23E-01	-3.68E+00	1.01E-03	0.3892703	-0.90715757	2.30E-01	1.50E-01	**	HPA	Weak/moderate staining
ILMN_3235113	TOMM6	-3.08E-01	-3.67E+00	1.02E-03	0.3892703	-0.91200447	NA	NA	***	HPA	Moderate staining
ILMN_3267017	RPLP1	-4.24E-01	-3.67E+00	1.02E-03	0.3892703	-0.9122183	1.60E-01	1.90E-01	*	HPA	No/w eak staining

Supplementary Table 3.

Gene	Direction of Primer	Sequence (5'-3')
RPLP2	Forward	AAGAAGATCTTGGACAGCGTGGGT
	Reverse	TACCCTGGGCAATGACGTCTTCAA
FAM129A	Forward	CCTTTCTTTGCTCAATGCAGGGCT
	Reverse	AGGGCACTGATGTTCATCCATCCA
PPA1	Forward	ATCTGGAACTATGGTGCCATCCCT
	Reverse	TTGGGTCATTGTCACCACAACAGC
DPM3	Forward	CTTCGGTTGAGTTTGGTCCGCTTT
	Reverse	CCAAAGCCACTGCGCTAATTTCGT
SOD2	Forward	AGATAGCTCTTCAGCCTGCACTGA
	Reverse	CACGTTTGATGGCTTCCAGCAACT
KRT7	Forward	TGTGGTGCTGAAGAAGGATGTGGA
	Reverse	TGTCAACTCCGTCTCATTGAGGGT
MME	Forward	AACAGCAACTTGGAGACGTTGTGC
	Reverse	CTCCAGCAAATGCTGCTTCCACAT

Supplementary Table 4.

Gene	Direction of Primer	Sequence (5'-3')
FAM129A	Forward	AGAAGGGAGCAAGGGCAAAT
	Reverse	AAAGGGCTCTGCTGGAAGTG
KRT7	Forward	AGGAGGGCAGTGAGATTCCT
	Reverse	GTATCGGCTGTGGAGCATGT
SOD2	Forward	TCTTTCCTGCGCTGTCTTGT
	Reverse	CCGCCAATGGCAGTGTAGAT
MME	Forward	ACCAAGTCTTACTGTCTAATTTTGC
	Reverse	TGACAGGTCTCAGTCACAACC
Chromosome 12B	Forward	AAAAAGCCAGGCACAGAAAA
	Reverse	AACCCCCAGCTTCTGGTAAT